The Dissertation Committee for Dae Kwan Ko certifies that this is the approved version of the following dissertation:

Clock-regulatory networks contribute to growth vigor in maize hybrids

Committee:

Z. Jeffrey Chen, Supervisor

Robert K. Jansen

Nigel Atkinson

Enamul Huq

Sibum Sung

Clock-regulatory networks contributes to growth vigor in maize hybrids

by

Dae Kwan Ko, B.S., M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2016

Acknowledgements

I would like to thank a number of people who have provided tremendous support for my Ph.D. study. First of all, my special thanks goes to my family, above all my supportive wife, Hyun Seung Lee. She has sacrificed many things for my dream of becoming a scientist. I deeply appreciate all of her supports.

I would like to thank my advisor Dr. Z. Jeffrey Chen for giving me the opportunity to be a part of his research team. He has patiently provided me invaluable knowledge and scientific lessons throughout the entire course of my graduate study. Without his support, advice and guidance, any of this work could not be finished. My appreciation extends to the committee members: Dr. Robert K. Jansen, Dr. Nigel Atkinson, Dr. Enamul Huq, and Dr. Sibum Sung, who offered me excellent comments and critiques of my research. I was very fortunate to work with both past and present members of the Chen lab, who provided great mentorship and friendship.

I would also like to thank collaborators for my Ph.D. study: Dr. Samuel H. Taylor for his contribution on photosynthesis measurement (Chapter 2), Dr. Frank G. Harmon and Dominica Rohozinski for their contribution on maize transgenic analysis (Chapter 3), and Dr. Qingxin Song for his contribution on ChIP-seq analysis (Chapter 4).

Clock-regulatory networks contribute to growth vigor in maize hybrids

Dae Kwan Ko, Ph.D.

The University of Texas at Austin, 2016

Supervisor: Z. Jeffrey Chen

Heterosis, or hybrid vigor, has been widely used in agriculture for more than a century. Despite extensive investigation and various models proposed, the molecular basis for heterosis remains largely elusive. In *Arabidopsis* interspecific and intraspecific hybrids, increased photosynthetic and metabolic activities are linked to altered expression of central circadian clock regulators, including *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*). It is unknown whether a similar mechanism mediates maize heterosis. In this dissertation, I investigated whether and how the circadian clock regulation contributes to growth heterosis in maize.

I reported that higher levels of carbon fixation and starch accumulation in maize hybrids are associated with altered temporal gene expression. Two maize *CCA1* homologs, *ZmCCA1a* and *ZmCCA1b*, are diurnally up-regulated in the hybrids. In *Arabidopsis ZmCCA1* complements the *cca1* mutant phenotype, and overexpressing *ZmCCA1b* disrupts circadian rhythms and heterosis. Furthermore, overexpressing *ZmCCA1b* in maize reduced chlorophyll content and plant height. Reduced height stems from reduced node elongation but not total node number in both greenhouse and field

conditions. Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) analysis revealed a temporal shift of ZmCCA1-binding targets to the early morning in the hybrids, suggesting that activation of morning-phased genes in the hybrids promotes photosynthesis and growth vigor. This temporal shift of ZmCCA1-binding targets correlated with nonadditive and additive gene expression in early and late stages of seedling development. These results could guide breeding better hybrid crops to meet the growing demand in food and bioenergy.

Table of Contents

List of Tables

List of Figures

Chapter 1. Background and Introduction

Heterosis, or hybrid vigor, is a fundamental biological phenomenon in which heterozygous hybrids show superior performance relative to their parents, in terms of growth, yield or resistance to environmental challenges. Heterosis has been widely applied to crop breeding and revolutionized the production [1-4]. Despite decades of extensive study, the molecular basis for heterosis is poorly understood. In this chapter, I review current understanding of heterosis and the circadian clock with three aims: (i) to summarize recognitions and applications of heterosis, (ii) to introduce the suggested link of heterosis with the circadian clock, (iii) to discuss the association of the circadian clock with growth and fitness in plants. The information guides my directions on this dissertation study.

1.1 Historical views of heterosis

A filial (F1) hybrid often outperforms its parents in growth and fitness such as increased size, yield and tolerance to environmental stresses. This phenomenon, known as heterosis or hybrid vigor, has been observed in a large number of different species including fungi [5], microorganisms [6], bivalve molluscs [7], animals [8-11] and plants [12-14]. Although this chapter focuses on heterosis in plants especially maize (*Zea mays*), readers are strongly encouraged to read published reviews on heterosis in other species [15].

The scientific discovery of heterosis dates back to Charles Darwin who first reported growth vigor of hybrid plants by analyzing growth patterns of cross-pollinated plants relative to that of self-fertilized plants in a systematic manner and concluded that cross-fertilization was generally beneficial; moreover self-fertilization was often deleterious (later known as inbreeding depression) [14]. Early in the twentieth century, George H. Shull and Edward M. East independently studied genetic basis of heterosis at different institutes, rediscovering it in maize breeding [16, 17]. Shull, who introduced the word "heterosis", suggested that self-fertilized maize plants, which are naturally outcrossed, have relatively reduced growth and yield [16]. This suggestion was consistent with the result of maize inbreeding experiments performed by East, which indicated the harmfulness of the low seed yield in inbred lines for hybrid production [17]. East further concluded that the degree of heterosis increases with genetic distance between the parents and interspecific hybrids show greater heterosis than intraspecific hybrids, which has commonly been accepted in the species [18-20].

As Shull's suggestion of crossing inbred lines was rapidly spreading in public and private research sectors in the 1920s, maize breeders were attracted to the economical advantages of heterosis for maize production while they had failed with other breeding methods [13], yet no one knew the genetic mechanisms of heterosis. Since the performance of the early maize inbred lines was not good enough to produce high yield of F1 hybrids, a double-cross hybrid was soon introduced in the hybrid breeding, in which two single crosses are crossed, generating a double cross; the idea was initially proposed by D. F. Jones in 1918 [21].

It was necessary to find an effective way to deliver hybrid seeds to farmers or to let farmers produce hybrid seeds. Since delivering hybrids seeds by "corn seed companies" was the best way for farmers to conveniently use hybrid seeds, the seed industry business started in 1920s and grew rapidly [1]. The first company to develop, produce and market hybrid corn seeds was the Hi-Bred Corn Company, later known as Pioneer Hi-bred International, Inc. In 1960s, artificially selected inbred lines were introduced and able to produce single-cross hybrids whose performance often surpassed double-cross hybrids. Thus, single-cross of high-yielding inbred lines allowed time-efficient and less-expensive hybrid production with no need of painstaking double-crossing. Maize breeders in public and private sectors have continuously provided superior inbred lines and hybrid combinations year after year. Since then, the yield of maize gradually increased in United Stated; an average of 11 tons hectare^{-1} in 2013 compared to 4.3 ton hectare^{-1} in 1961 [22] (Figure 1.1). Aided by the use of synthetic nitrogen and better weed control, heterosis seems to be a major factor for the increase. Wilbert A. Russell reported that 60% or more of the increase from 1930 to 1970 is due to genetic efforts including selecting superior inbred lines and hybrid combinations [23]. Now, ~95% of maize acreage in United States is planted with hybrids. In addition to maize, heterosis has been a major contributor in rice, sorghum, sunflower, tomato and cotton breeding, increasing the global productivity [2, 24-26].

1.2 Proposed models for heterosis

Since the breeding of most crops and vegetables is largely based on the principle of heterosis [2, 24-26], better understanding of those principles would enhance crop yield and performance, which must double to feed 9 billion people estimated by 2050 [27-29]. However, despite decades of intensive work, the mechanism for heterosis still remains elusive. Several genetic models have been proposed to explain heterosis [30-32]. The first of these is the dominance model in which slightly deleterious recessive alleles from one parent are complemented by dominant alleles from the other parent in F1 hybrids (Figure 1.2A). As a result, hybrids have a better performance than the parents. Additive gene expression in the hybrids, which is equal to the mid-parent value (MPV), is responsible for this action. The second is the overdominance model, which is about allelic interaction at single heterozygous locus (Figure 1.2B). According to this model, F1 hybrids carry both alleles, which are not in dominant and recessive relationship but function synergistically to promote heterosis. Nonadditive gene expression in the hybrids, which represents deviation of gene expression in the hybrid from MPV, is responsible for the action. The pseudo-overdominance is a genetic intermediate between the two models above; a simple case of dominance complementation, in which the two recessive mutations are genetically linked in repulsion (Figure 1.2C). Although there have been numerous amount of studies in plants and other species favoring one model or another over decades, these traditional models have not clearly explained the complexity of heterosis. For example, in hybrid rice, heterosis is found to be linked with dominance, overdominance, pseudo-overdominance and all of the above [33] so is it in hybrid maize [34]. It has been suggested that these traditional models should be updated with gene

expression and regulatory networks that are often altered in the hybrids relative to the parents [3, 25, 31].

Consistent with this concept, recent studies have discovered a link between altered circadian clock regulation and increased levels of photosynthetic and metabolic activities and biomass in *Arabidopsis* allotetraploids (interspecific hybrids) [35] and *A. thaliana* intraspecific hybrids [36, 37]. These results collectively indicate that the expression of central circadian clock genes is epigenetically altered in hybrids, which in turn increases expression levels of downstream genes involved in energy and metabolic pathways, promoting carbohydrate metabolism during the day and night. The more starch accumulates during the day; the more starch can be degraded at night to stimulate growth, leading to biomass heterosis [31].

1.3 The circadian clock in plants

As the 24 h rotation of the earth on its axis leads to environmental changes of light and temperatures that predictably generate the day-night cycle, virtually all living organisms (Bacteria, Archaea and Eukaryota) [38] have evolved to possess an internal timekeeping mechanism, called the circadian clock, which allows organisms to coordinate the time of the day and adjust their biological processes to the 24 h cycle. Thus, circadian clocks enable them not simply respond to the environmental change but also anticipate and prepare for the next coming change. The first scientific evidence was documented as early as the eighteenth century. In 1729, the French astronomer De Mairan reported the first experimental evidence of this characteristic that the daily

rhythms in leaf movement of a heliotrope plant, most likely *Minosa pudica*, persisted under constant dark condition, indicating that the rhythms are from an endogenous origin [39, 40]. Since then, an enormous amount of evidence began accumulating that circadian clocks are pervasive in archaea, fungi, animals and other plants [41-43]. Importantly, there are fundamental characteristics for circadian clocks [44-46]. First, as the term circadian, coined by Franz Halberg in 1959, was originated from the Latin words "circa" (about) and "dies" (day), circadian clocks generate the subset of biological rhythms with period of ~24 hours. Second, circadian clocks are endogenous and self-sustaining, so they persist in the absence of the environmental inputs, typically under constant light, dark or temperature. Third, circadian clocks are under temperature compensation: the period of circadian rhythms is constantly maintained over a fluctuation of ambient temperatures [47]. This characteristic might act as a buffering mechanism against changes in cellular metabolisms. Collectively, these characteristics are used to define circadian clocks.

The advance of genetic and molecular techniques has allowed the molecular components of the circadian clock to be identified in model organisms, including flies, mice and *Arabidopsis*, to date [42, 43, 48]. Despite the shared molecular architecture of multiple transcription-translation feedback loops in the circadian clock, each species possesses specific clock components and biochemical processes. In *Arabidopsis* and plants in general, the circadian clock consists of a central, morning and evening feedback loops, compromising a gene network of circadian oscillators (Figure 1.3) [43, 45, 49]. The central feedback loop is based on the two partially redundant morning-expressed MYB transcription factors, CIRCADIAN CLOCL-ASSOCIATED 1 (CCA1) and LATE

ELONGATED HYPOCOTYL (LHY), which directly repress transcription of an evening-expressed pseudo-response regulator (PRR), TIMING OF CAB EXPRESSION 1 (TOC1). TOC1, interacting with CCA1 HIKING EXPEDITION (CHE), directly represses *CCA1* and *LHY* expression. In the morning feedback loop, CCA1 and LHY activate expression of three additional *PRR* genes (*PRR5*, *7* and *9*), which in turn repress *CCA1* and *LHY* expression. Therefore, these PRR proteins, including TOC1, let *CCA1* and *LHY* expressed only during a limited time of each day. The evening feedback loop consists of TOC1 and GIGANTEA (GI), as part of a hypothetical component 'Y', and Evening Complex (EC), which consists of LUX ARRHYTHMO (LUX), EARLY FLOWERING 3 (ELF3) and ELF4. They are in double negative feedbacks with the components of central (CCA1 and LHY) and morning feedback loops (PRR7 and PRR9). The core clock, composed of these interlocking feedback loops, coordinates numerous output biological events with appropriate time of day, primarily through transcriptional regulatory networks [31, 43]. However, despite the great interest and importance, less information is available on the gene regulatory networks linking the core clock with the rhythmic biological events.

1.4 Effects of the circadian clock on plant growth and fitness

The circadian clock widely regulates plant physiology including growth and fitness with the daily and seasonal cycles. In *Arabidopsis*, when the endogenous clock matches the external diurnal cycle, CO_2 fixation, photosynthetic activities and fitness are increased [50]. Disrupting the clock functions causes disadvantages on growth and

fitness. For example, a *CCA1* overexpression mutant (*CCA1-ox*) lacks circadian rhythms and, as a result, displays reduced photosynthetic growth and fitness [51]. The double mutant *cca1 lhy*, having a circadian period shorter than 24 h, accumulates less starch and is unable to properly set the rate of starch degradation to match the length of night [52]. In this sense, it is hypothesized that optimized clock activity will enhance productivity of crops such as maize [53, 54].

Maize performs C4 photosynthesis, which anatomically and biochemically distinct from C3 photosynthesis in *Arabidopsis* [55]. Photosynthetic activities in maize are subject to diurnal regulation [56]. In particular, sucrose accumulation rate increase during the day and peaks at 15:00 when the rate of starch accumulation increases until dusk. Starch mobilization occurs at the beginning of dusk, and all of the starch was depleted by the end of the night. Mobilization of starch at night likely promotes growth, which, after temperature correction, is greatest at night [57]. This diurnal regulation of carbohydrate metabolism is consistent with diurnal expression of genes in maize leaves. Approximately 10% of ~13,000 transcripts examined display circadian expression patterns [54]. The majority of cycling genes peak their expression at subjective down and dusk, similar to other plant circadian systems. These genes participate in fundamental physiological processes, including photosynthesis, carbohydrate metabolism, cell wall biogenesis, and phytohormone biosynthesis pathways. In another study, 23% of expressed transcripts exhibit a diurnal cycling pattern in leaves [58]. By constrast, in developing ears only core circadian clock oscillators plus some fewer gene transcripts are diurnally regulated. This suggests tissue-specific circadian regulation in maize as

observed in *Arabidopsis*, where the clock in roots is a slave version of the clock in shoots [59], and this clock is established early during embryo development [36]. Furthermore, developmental shifts in gene expression can also influence biomass accumulation. In sorghum, SbPRR37 activates expression of several downstream genes that repress flowering in long days [60]. *SbPRR37* expression is dependent on light and regulated by the circadian clock. In short days, *SbPRR37* is not expressed during the evening-phase, allowing sorghum to flower. This suggests that a mutation in the clock-interacting gene can change flowering time and promote biomass production in sorghum.

Quantitative trait loci (QTLs) for agronomic traits in crop plants are likely associated with the clock pathway [61]. For example, yield-related QTLs are mapped to clock-related and light signaling genes in the super-hybrid rice [62]. Flowering QTLs are associated with circadian clock regulators, *GIGANTEA* in soybean [63] and *CONSTANS* in sorghum [64]. A recent study finds that allelic variations at clock loci in tomato varieties are linked to its domestication [65]. In maize, agronomic traits are controlled by many small-effect QTLs [66], and several clock regulator homologs are identified as *a priori* candidate genes among genome-wide association QTLs [67]. These studies support the notion that the circadian clock mediates flowering, growth, and heterosis in crop plants including maize as in *Arabidopsis* plants [4]. Collectively, These studies support the notion that the circadian clock could be one of factors that affect crop productivity, biomass or heterosis.

1.5 Perspectives

Despite extensive efforts of the research community to identify the molecular basis of heterosis using a number of different approaches, it is still far from the complete understanding. Beyond the traditional “dominance” and “overdominance” models, heterosis can rather be considered as consequences of genomic interactions at multiple molecular and network levels including genomic, epigenomic, transcriptomic, proteomic and metabolomic levels. Especially, many of gene expression variations could be associated with a few major network regulators, which simultaneously regulate thousands of downstream genes associated with many metabolic and physiological pathways, leading to heterosis [31]. One of such regulators could be the circadian clock. Now, given the complex nature of heterosis, integrative approaches of genetics, plant physiology, biochemistry and genomics are highly desirable to identify the link between heterosis and the circadian clock in crops such as maize. These approaches should allow researchers to address whether and how the circadian clock contributes to growth vigor in hybrid crops.

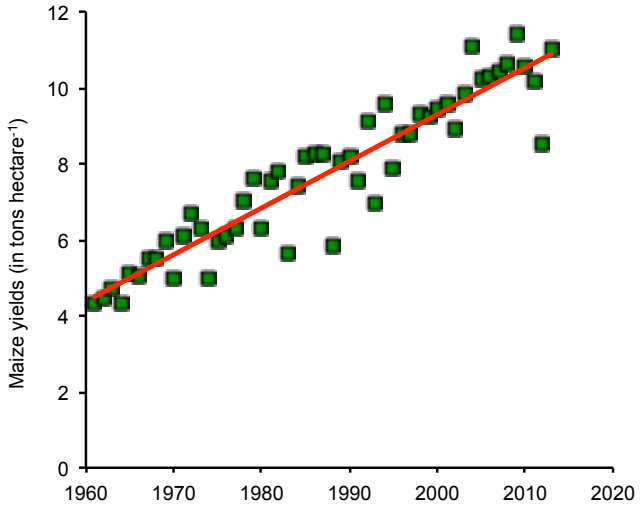

Figure 1.1 Annual average yields of maize in United States.

Annual average yields of maize (in tons hectare^{-1}) for each year from 1961 to 2013, according to database in Food and Agriculture Organization of the United Nations, 2015. The solid red line is the linear regression on this data.

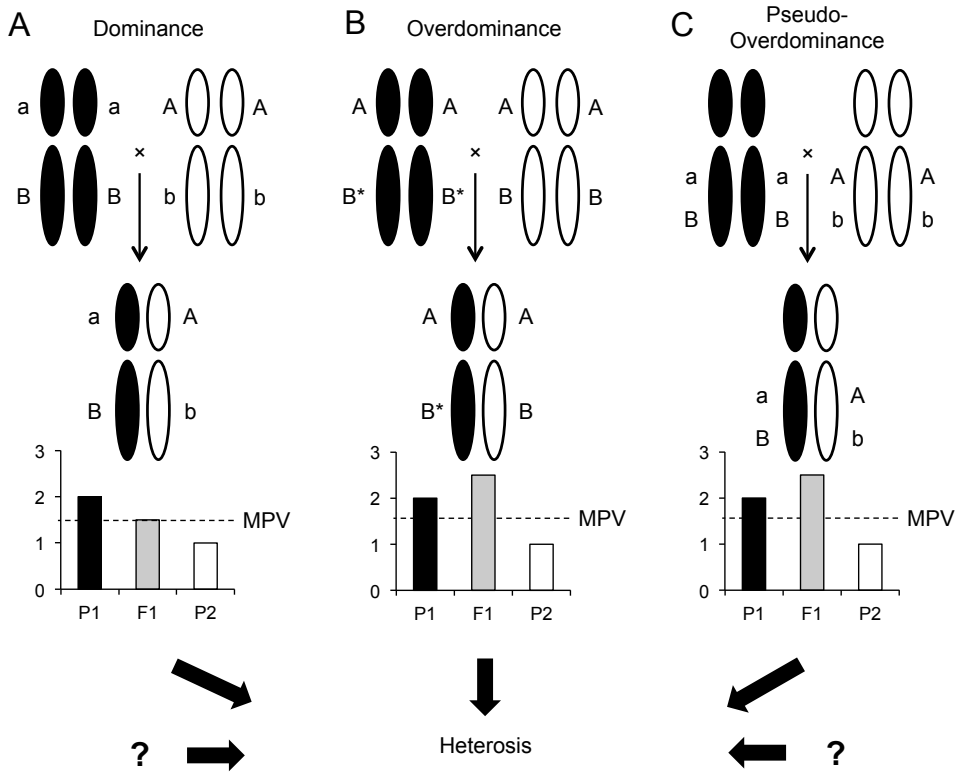

Figure 1.2 Proposed genetic models and gene expression levels for heterosis.

Multiple linked or unlinked genes control heterosis. “A” and “B” represent superior alleles whereas “a” and “b” represent slightly deleterious alleles. For the overdominance model (B), a dominant-recessive relationship is not considered as shown “B*” and “B”. Black and white colors represent parent 1 (P1) and parent 2 (P2), respectively. The histogram below each model shows expected gene expression levels.

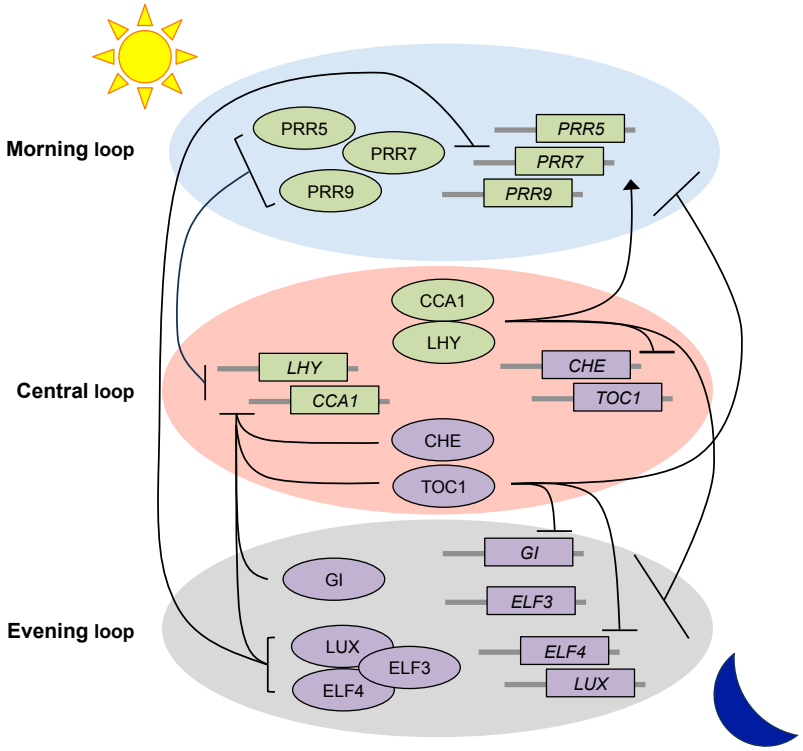

Figure 1.3 A simplified model for gene networks of the circadian clock in *Arabidopsis*.

The circadian clock oscillator in *Arabidopsis thaliana* comprises a series of interlocked feedback loops: morning, central and evening loops indicated by blue, red and gray circle areas, respectively. Each curved black line represents transcriptional regulation: either repression or activation indicated by arrow or horizontal line, respectively. Morning-phased genes (boxes) or proteins (circles) are indicated by green while evening-phased genes (boxes) or proteins (circles) are indicated by violet.

Chapter 2. Physiological basis for growth heterosis in maize hybrid seedlings

2.1 Introduction

Heterosis has been documented in many plant species since the early scientific description by Charles Darwin [14]. For example, rice hybrids within a species or between the subspecies *Oryza sativa* ssp. *indica* and *O. sativa* ssp. *japonica* show increased growth rate and stress tolerance compared to the parents [68]. Hybrid rice produces grain yields 30-40% more than does conventional rice [69]. Since the first release of the commercial hybrid rice variety in 1976, hybrid rice has been planted in 50% of the total rice area in the world [70, 71]. Cotton allotetraploids (*Gossypium hirsutum*, genome AADD) were formed by hybridization between extant diploid progenitors approximately 1-2 million years ago [72]. *G. hirustum* species, which is charged for 95% of cotton production worldwide, yields more abundant and higher quality fibers than the progenitors [73]. Hexaploid bread wheat (*Triticum aestivum L.*, genome BBAADD) was hybridized between a domesticated tetraploid wheat (*T. turgidum*, genome BBAA) and a diploid goat grass (*Aegilops tauschii*, genome DD) nearly 10,000 years ago [74, 75]. Owing to its greater physiological and ecological plasticity than its progenitors [74, 76], the hexaploid bread wheat has successfully been domesticated and spread throughout the globe. Now it feeds roughly one third of the world population [22].

Maize, which is one of the world's most important crops, has been an excellent material for both the application and research not only because of its relevance to

agriculture, but also because it shows the most evident heterosis [1, 13] (See Chapter 1). Hybrid maize, accounting virtually all grown corns worldwide [34], generally displays vigor during vegetative growth such as biomass, increased productivity including grain yield and greater stress tolerance to environmental stresses relative to the inbred parents [77-79]. Although adult traits have extensively been investigated for heterosis, growing evidence has suggested that growth heterosis is already manifested during early stages of development in maize and other plant species [36, 80-82]. Thus, understanding the physiological basis for growth heterosis at early stages of development would provide an opportunity to identify a clue for the underlying molecular mechanisms of heterosis. Nevertheless, the essential physiological characterization is currently missing.

Here, to address this issue, I analyzed multiple growth-related traits in maize inbred and F1 hybrid seedlings and identify physiological basis for growth heterosis. These data indicate that growth heterosis is established in the early seedling stage without an initial advantage of seed size and mass and subsequently maintained during seedling development. The measurement of diurnal photosynthetic rate and carbohydrate content reveal that the hybrids have increased capacity of carbon fixation. These analyses suggest a role for diurnal regulation of photosynthesis and carbon metabolism in growth heterosis.

2.2. Materials and methods

Plant materials and growth conditions

The inbred lines B73 and Mo17 and their reciprocal F1 hybrids B73XMo17 (BM) and Mo17XB73 (MB) were used for this study. All maize plants were grown at 600 μmol m^{-2} s^{-1} under 16L:8D (hours of light:dark) cycle with temperature 28 °C (light) and 23 °C (dark) except for the experiments to measure photosynthetic rates, starch and sucrose content. The experiments were carried out in a glasshouse where natural sunlight (maximum photosynthetically active radiation 598 ± 354 μmol m^{-2} s^{-1}, mean ± SD) was supplemented with fluorescent lighting to achieve a 16L:8D cycle with daily maximum temperature at 26.7 ± 2.1 °C (mean ± SD), daily minimum temperature at 19.4 ± 0.4 °C (mean ± SD), relative humidity at 57 ± 7% (mean ± SD), and net CO_2 exchange examined during the day. The plants were placed in a randomized design and rotated on a daily basis to minimize positional effects.

Growth trait measurement

The growth traits included seed mass, aerial biomass, plant height, and leaf-blade length and area. For seed weight, 10 seeds randomly selected from three replicates for each genotype were weighed. For aerial biomass, aboveground seedlings for each genotype were collected at 3, 4, 5 and 6 days after planting (DAP) (n = 4-8). The biomass was weighed after desiccation for 48 h at 80 °C. Plant height and leaf-blade length for each genotype were measured daily from 5 DAP to 12 DAP (n = 5). Plant height was measured as the distance from the ground to the top of seedlings. Leaf-blade length was measured as the distance from the base to the tip of the 2^{nd} leaf. The 2^{nd} leaf at 12 DAP for each genotype was photographed and processed for leaf area using NIH ImageJ

software (http://imagej.nih.gov/ij/) [83]. All experiments were replicated three times, unless noted otherwise.

Photosynthesis, starch and sucrose measurements

Photosynthetic rate was measured on fully expanded 2nd leaves of seedlings at 12 DAP (n = 9), grown under 16L:8D cycle with natural sunlight, every 2 hours from dawn (ZT0) to dusk (ZT14). A LI-6400XT portable photosynthesis analyzer (LICOR Environmental, Lincoln, NE) equipped with a light source (LI-6400-40) and CO_2 mixer (LI-6400-01) was used to determine net CO_2 assimilation at a reference CO_2 concentration of 400 μmol mol^{-1}, with cuvette conditions set to match ambient conditions at the start of each measurement period. The experiment was reproduced twice, in February and April 2012, respectively. Starch and sucrose contents were measured on source (middle-tip of the 2nd leaf) and sink (base of the 2nd leaf) of additional replicate plants grown alongside those used for photosynthesis assays. Pooled plants in three biological replicates at ZT14 were used for testing. Leaf discs were collected using an 11 mm-diameter cork borer, weighed and immediately frozen in liquid nitrogen. The frozen tissues were ground, mixed with a homogenization buffer (500 mM MOPS pH 7.5, 5 mM EDTA, 10% ethyl glycol), and then filtered through Miracloth (CalBiochem, San Diego, CA). After centrifugation, pellets were dissolved in DMSO to extract the insoluble carbohydrate fraction, while supernatant was transferred to a new tube as the soluble carbohydrate fraction. The starch content was measured from the insoluble carbohydrate fraction using a commercial assay kit according to the manufacturer's instruction (R-Biopharm, Darmstadt, Germany). Sucrose content was measured from the soluble

carbohydrate fraction using a commercial assay kit according to the manufacturer's instruction (K-SUFRG Megazyme, Bray, Ireland) and as previously described [35].

2.3 Results

Early establishment of growth vigor in maize hybrids

F1 hybrid seeds in maize are often larger than those of the parents, as observed in *Arabidopsis* [37, 80, 81]. Seedling growth vigor was reported to correlate with larger seeds in F1 hybrids [17]. This hypothesis was not supported by seed size analysis: seed weight was similar in maize inbred lines (B73 and Mo17) and their reciprocal F1 hybrids (B73XMo17, BM, and Mo17XB73, MB) (by convention, the female parent is listed first in a genetic cross) (Figure 2.1A). The data indicates that seedling growth heterosis in maize may occur after germination without an initial advantage in seed size and mass. Yield heterosis in these hybrids is likely associated with larger and longer cobs that produce many more seeds than those in the inbred lines.

To identify the developmental phases at which growth heterosis is established in maize, diverse growth related phenotypes were daily monitored in the inbreds and hybrid seedlings. In spite of similar seed size, seedling growth heterosis occurred soon after germination (Figure 2.1B), and leaf area of the hybrids was significantly greater in F1 hybrids than in the parents (Figure 2.1C). F1 hybrids showed biomass heterosis 5 days after planting (5 DAP) (Figure 2.1D), and leaf length of the hybrids was significantly greater than the best parent from the emergence of the first true leaf 5 DAP until 14 DAP (Figure 2.1E). Biomass and plant height heterosis was significantly different between the

reciprocal crosses, and vigor was higher when Mo17 was used as the maternal parent. However, this difference in height decreased and was no longer obvious after two weeks (Figure 2.1D and Figure 2.2). These data suggest early establishment of seedling growth heterosis and subsequent maintenance of this heterosis during seedling development.

The degree of heterosis is often defined by either mid-parent heterosis (the performance of a hybrid relative to the average of its two parents; MPH) or better-parent heterosis (the performance of a hybrid relative to the parent showing the better performance; BPH) [32]. To identify the degree of heterosis during seedling growth, the percent BPH was calculated at 5 DAP, 8 DAP, 11 DAP and 14 DAP (Figure 2.2). The percent BPH at the early stages (5 DAP to 8 DAP) was not remarkably different from one at the later stages (11 DAP to 14 DAP). Taken together, the data suggest an establishment of seedling growth heterosis and subsequent maintenance of this heterosis during seedling development.

Enhanced capacity of carbon fixation in the hybrids

Seedling growth rates were associated with increased levels of photosynthetic activities including CO_2 assimilation, which is under diurnal regulation as observed in maize inbred lines [56]. Thus, it was tempting to speculate that the growth heterosis is also associated with increased levels of photosynthetic activities in the hybrids. To address this hypothesis, CO_2 assimilation rate was measured every 2-hours from zeitgeber time 0 (ZT0 = dawn) to ZT14 under 16L:8D cycle. Positive net CO_2 assimilation increased from ZT0, peaked at ZT8, and then declined through the remaining light period, having effectively ceased at ZT14 (Figure 2.2A). Notably, the amplitude of

CO_2 assimilation significantly increased in both hybrids relative to the mid-parent value (MPV), especially during the morning phases. This increase in photosynthetic rate of the hybrids must be added to the expected effect of greater leaf area on whole plant carbon gain relative to the inbreds.

The increased photosynthetic activities promote starch synthesis and mobilization. In a maize leaf, the middle-tip region represents a source where carbohydrates are synthesized, whereas the base region represents a sink where carbohydrates are stored, utilized in growth, and mobilized [55, 84]. In the source region, starch amount was significantly more in the hybrids than the MPV, but the sucrose content was similar (Figure 2.2B). In the sink region, both starch and sucrose accumulated significantly more in the hybrids than the MPV. Little photosynthesis activity is present in the sink region of a maize leaf [85]; thus starch accumulation is likely caused by the increased CO_2 assimilation in the source region, with more carbohydrates being mobilized to the sink region. More sucrose synthesis and mobilization in the hybrids provides more carbohydrates for seedling growth, suggesting a role for the diurnally regulated enhancement of carbon fixation capacity in growth heterosis.

2.4 Discussion

In this Chapter, I analyzed multiple growth-related traits in the maize inbred and F1 hybrid seedlings and observed the early establishment of growth vigor, followed by subsequent maintenance, in the hybrids. The analysis provided evidence that the hybrids possess increased CO_2 assimilation rate and carbohydrate content relative to the MPV,

which indicates that the increased carbon fixation capacity contributes to growth vigor in the hybrids relative to the inbreds.

These data revealed several points into the overview of growth heterosis in maize seedlings. First, growth heterosis in maize hybrids occurs after germination without an initial advantage in seed size and mass (Figure 2.1). No significant differences of seed mass and size were observed among genotypes in this study, rather heterosis for the growth traits analyzed constantly started being observed soon after germination. In support of our data, in the 12 possible reciprocal hybrids of maize inbred lines UH002, UH005, UH250 and UH301, heterosis for primary root length was established at 3 days after germination [82], which is equal to 5 DAP; seeds mostly germinate at 2 DAP. In addition, the hybrids constantly showed growth vigor without prominent increases of BPH from 5 DAP to 14 DAP (Figure 2.2). Therefore, although growth traits examined in this study do not represent all growth heterotic traits, it is evident that once established at the early stage of development, growth heterosis is subsequently maintained during the growth.

Second, the hybrids have not only larger photosynthetic areas but also higher net photosynthesis rate than the inbreds (Figure 2.3A). Consistent with this observation, maize hybrids one-way crossed from eight inbred lines showed extensive heterosis for net photosynthetic rate [86]. Crosses of two distinct accessions in *Arabidopsis thaliana* also showed increased photosynthetic efficiency compared to the parents [87]. In our data, notably, the increased efficiency in the hybrids was observed only in the morning phase (ZT0, ZT4 and ZT6), indicating that there might be time-specific mechanisms in

heterosis for net photosynthetic rate, which is under circadian regulation [50]. However, since there are different mechanisms that contribute to net photosynthetic rates in crops (e.g. C4; maize and sorghum, C3; rice and wheat) [88], additional experiments are needed to further demonstrate the prevalence of heterosis for net photosynthetic rate in other hybrids of maize or other crops.

Third, the two different regions of the leaf showed unique patterns of heterosis for carbohydrate contents (Figure 2.3B). The analyses of carbohydrates on source and sink regions of the mature leaf showed that the hybrids have higher starch contents both in source and sink regions while having higher sucrose only in sink region relative to the inbreds. This indicates that heterosis for carbon fixation capacity should be taken into account separately in source and sink regions of a leaf in maize hybrids, unlike *Arabidopsis* hybrids, indicating another layer of complexity in the biology. Namely, the increased capacity of carbon fixation in the hybrids may be the result of a cumulative effect from photosynthesis to starch and sucrose metabolisms, which consequently contributes to the growth vigor in the hybrids.

Taken together, these results demonstrate that growth heterosis in maize hybrids is established at early seedling stages with increased capacity of carbon fixation. In addition, increased capacity of photosynthetic activities and carbohydrate contents in the hybrids, which are under diurnal and circadian regulations, suggests a role of the circadian clock in growth heterosis in maize as in *Arabidopsis* [31].

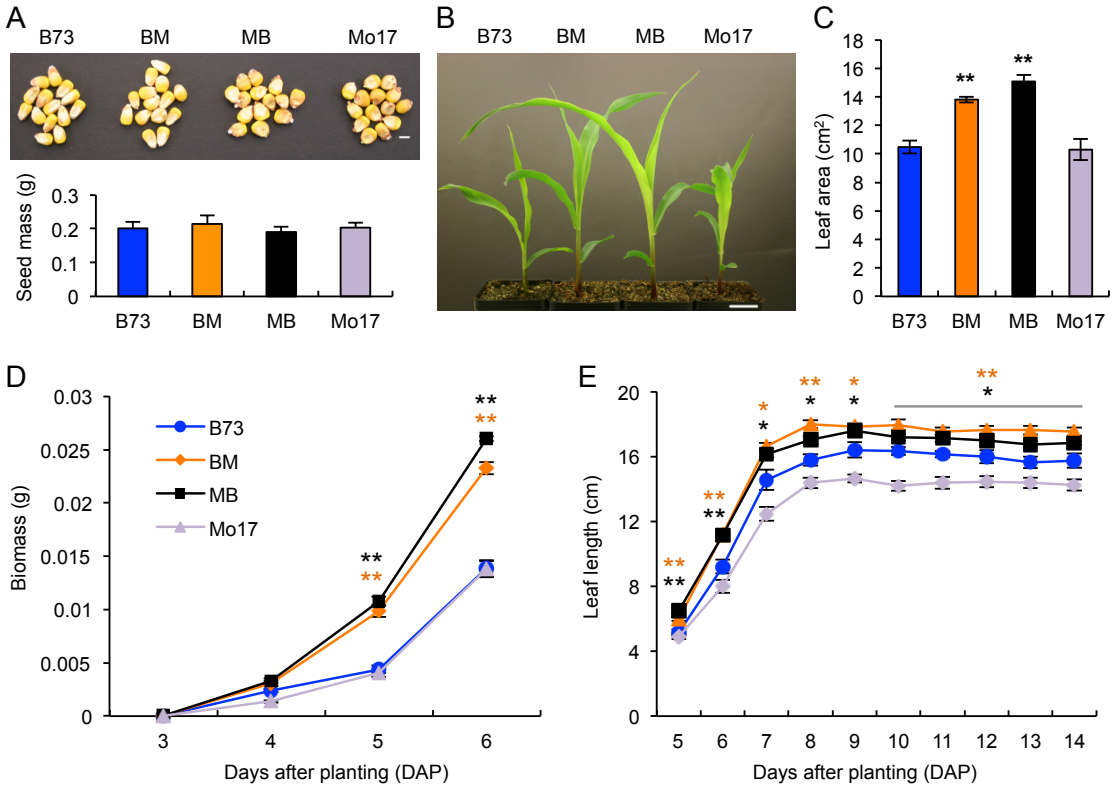

Figure 2.1 Growth heterosis starts at the early seedling stage.

(A) Seed morphology (upper panel) and weight (lower panel, means ± SEM) in maize inbred lines (B73 and Mo17) and their reciprocal F1 hybrids (B73XMo17, BM and Mo17XB73, MB). Scale bar = 10 mm. (B) Seedling phenotypes showing growth vigor in maize F1 hybrids. Scale bar = 10 mm. (C) Leaf area of F1 hybrids relative to the inbreds (means ± SEM, n = 10). (D) Growth curve of biomass in the maize inbreds and F1 hybrids (means ± SEM, n = 6). (E) Growth curves of leaf length in the maize inbreds and F1 hybrids (means ± SEM, n = 5). Heterosis for biomass and leaf length was observed at 5 DAP in the hybrids. Note that the earliest time to measure biomass was 3 DAP when young leaves emerged. Significant difference between hybrids and the best parent was calculated using Student's t-test, *p-value < 0.05 and **p-value < 0.01.

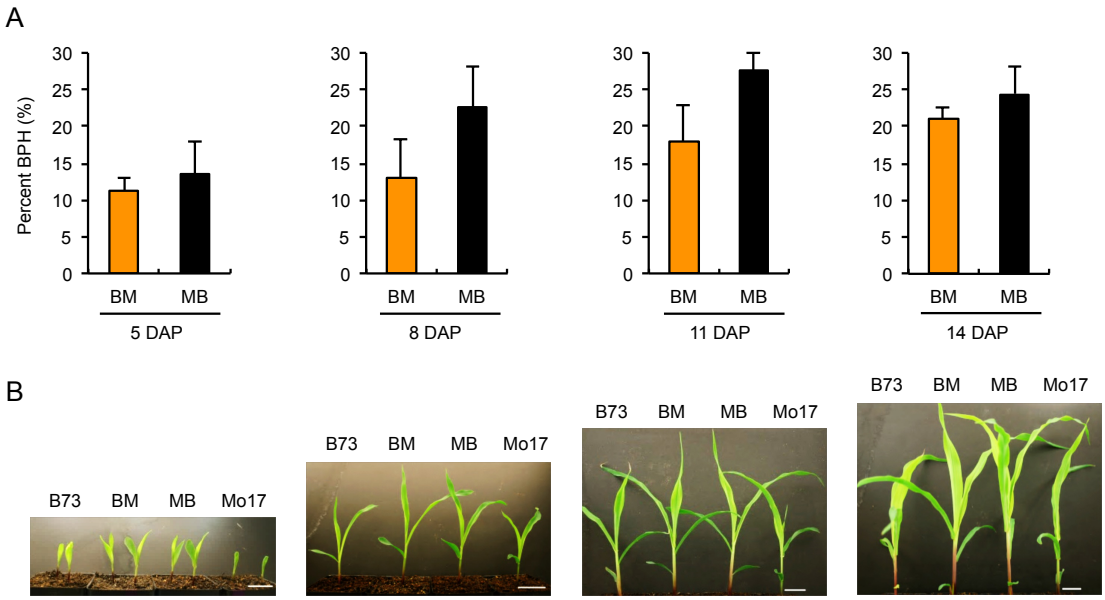

Figure 2.2 The early-established heterosis is subsequently maintained.

(A) The percent better- parent heterosis (BPH) is shown for plant height. The percentage BPH (means ± SEM) was calculated for each biological replicate (n = 5) as: %BPH = ((hybrid – better-parent)/better-parent) X 100. (B) Representative growth vigor in the reciprocal hybrids at different developmental stages; 5 DAP, 8 DAP, 11 DAP and 14 DAP. Scale bar, 40 mm.

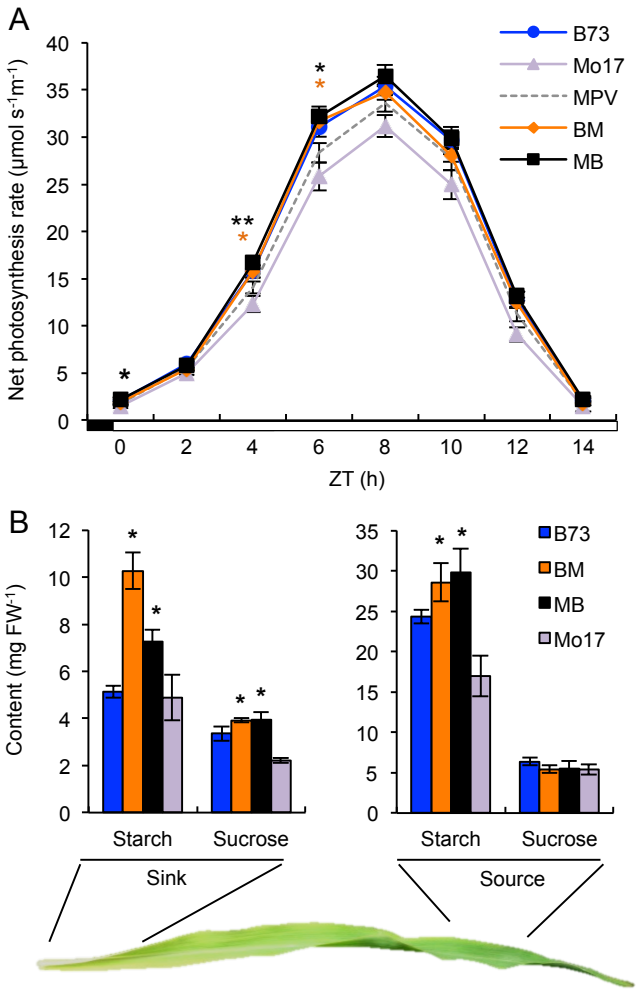

Figure 2.3 Increased capacity of carbon fixation in the F1 hybrids relative to the inbreds.

(A) Diurnal rhythms of net photosynthesis rate in the maize inbreds and hybrids (means ± SEM, n = 9). (B) Starch and sucrose accumulation in sink and source regions of a mature leaf (means ± SEM, n = 3) at ZT14. FW, fresh weight. Significant difference between hybrids and MPV was calculated using Student's t-test, *p-value < 0.05 and **p-value < 0.01.

Collaborated with Dr. Samuel H. Taylor for the photosynthesis measurement.

Chapter 3. Molecular and circadian characterization of maize *CCA1* homologs

3.1 Introduction

As a central component of the circadian clock, *CCA1*, a MYB DNA-binding domain transcription factor gene, mediates physiological and metabolic aspects involved in plant growth and fitness [31, 40]. For example, *Arabidopsis CCA1-ox* mutant lacking circadian rhythms displays reduced photosynthetic growth and increased flowering time [50, 51, 89]. The *cca1 lhy* double mutant, which has a circadian period shorter than 24 h, accumulates less starch and is unable to properly set the rate of starch accumulation to match the length of night [52]. Thus, an optimized activity of central clock regulators, such as CCA1, has been thought to potentially contribute to growth heterosis. Indeed, recent studies showed that altered expression of *CCA1* and *LHY* was linked to increased photosynthetic and starch metabolism activities in *Arabidopsis* allotetraploids and intraspecific hybrids [35]. However, whether maize or other crops utilize their clock in the similar way remains unknown.

The maize genome contains two *CCA1* homologs, now designated *ZmCCA1a* and *ZmCCA1b*. Protein sequences encoded by *ZmCCA1a* and *ZmCCA1b* are highly similar. Two paralogous genes are rhythmically expressed in a similar fashion with higher expression of *ZmCCA1b* [58, 90]. Despite this knowledge, circadian and molecular functions of the maize homologs are largely uncharacterized.

To better understand the circadian clock and its relation with growth heterosis in maize, I here characterize molecular and circadian functions of the two maize *CCA1*

homologs. The clock regulator genes, including *ZmCCA1a* and *ZmCCA1b*, are nonadditively expressed in maize hybrids, consistent with the finding in *Arabidopsis* allotetraploids [35] and intraspecific hybrids [36, 91]. The recombinant ZmCCA1b protein, but the truncated ZmCCA1a lacking MYB DNA-binding domain, is able to bind to promoters of other clock regulators specifically via EE or CBS *in vitro*. In *Arabidopsis* *ZmCCA1* complements the *cca1-11* mutant phenotype, and overexpressing *ZmCCA1b* disrupts circadian rhythms and growth vigor. Furthermore, overexpressing *ZmCCA1b* in maize reduced chlorophyll content and plant height. These findings uncover evolutionary conservation of circadian clock function between *Arabidopsis* and maize CCA1 homologs, and that DNA binding activities conferred by the MYB DNA-binding domain are required for the clock function.

3.2 Materials and methods

Gene expression analysis

For gene expression analysis, aerial tissues were harvested and immediately frozen in the liquid nitrogen. Total RNA was extracted from the frozen tissues by PureLink^{TM} Plant RNA reagent (Invitrogen, Carlsbad, CA) and treated with RQ1 DNaseI (Promega, Madison, WI) according to the manufacturers' instructions. For cDNA synthesis, 1 μg of DNaseI-treated total RNA was incubated with Omniscript reverse transcriptase (Qiagen, Valencia, CA) in the presence of 10 μM random hexamer (GeneLink, Hawthorne, NY). For qPCR, FastStart Universal SYBR Green Master (ROX) (Roche Applied Science, Indianapolis, IN) was used in the presence of gene-specific

primers and template cDNAs in an ABI7500 (Applied Biosystems, Foster City, CA) or LightCycler 96 machine (Roche Applied Science, Indianapolis, IN). The control was *18S rRNA* to estimate the relative expression levels of each gene in three biological replicates. A list of primers for gene expression analysis is provided in Table 3.1.

Purification of recombinant ZmCCA1b and ZmCCA1a proteins

The coding sequence (CDS) of *ZM2G474769* (*ZmCCA1a*) was amplified from B73 cDNA by the primer pair 5'-GAATTCATGCCCTTGAGCAATGAG-3' (*Eco*RI, underlined) and 5'-GTCGACTCATGTTGATGCTTCACTAT-3' (*Sal*I). The full-length *ZmCCA1b* (*ZM2G014902*) CDS was amplified from B73 cDNA by the primer pair 5'-GGATCCATGGAGGTGAATTCCTCTGGC-3' (*Bam*HI) and 5'-GTCGACTTATGTGGATGCTTCGCTATC-3' (*Sal*I). The *ZmCCA1a* or *ZmCCA1b* cDNA fragment was cloned into a pGEM-T (Promega, Madison, Wisconsin). After sequence verification, the *ZmCCA1a* or *ZmCCA1b* CDSs was subcloned into pMAL-C2 (New England BioLabs, Beverly, MA) through *Eco*RI/*Sal*I and *BamH*I/*Sal*I restriction sites, respectively. *E. coli* strain Rosetta-gami B competent cells (Novagen, Madison, WI) was used to transform empty pMAL (expressing maltose-binding protein, MBP), pMAL-ZmCCA1a or pMAL-ZmCCA1b, which were grown in 4 ml of Luria-Bertani (LB) media with Carbenicillin (100 mg/L) at 37 °C for 18 h. The overnight cultures of Rosetta-gami B cells containing pMAL, pMAL-ZmCCA1a or pMAL-ZmCCA1b construct were diluted into 1:100 in 80 ml LB media with Carbinicillin (100 mg/L) and grown at 37 °C to an OD_{600} value of 0.5, when isopropyl-β-D-thiogalactoside (IPTG) (0.1 mM) was added. After 20 h of additional incubation at 16 °C, cells were harvested after

centrifugation at 4,000 *g* at 4 °C for 10 min and resuspended in 2 ml of Column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA). After frozen at -20 °C for 18 h, cells were lysed by a Bioruptor Sonicator (Diagenode, Sparta, NJ) and centrifuged at 20,000 *g* at 4 °C for 20 min. The cleared cell lysates were diluted 1:5 with Column buffer, and loaded on amylose-coupled agarose resin columns prepared according to the manufacturer's instruction (New England BioLabs, Beverly, MA). After columns were washed with 12 volumes of Column buffer, MBP, rZmCCA1a and rZmCCA1b were eluted with elution buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM Maltose). After filtration by Amicon Ultra 100 K (Millipore, Darmstadt, Germany), the purified MBP, rZmCCA1a or rZmCCA1b was aliquoted and stored at -80 °C.

Electrophoretic mobility shift assays (EMSA)

DNA probes were generated by annealing PAGE-purified sense and antisense oligonucleotides (Table 3.2). The double-stranded oligonucleotides were [^{32}P] end-labeled using a T4 Polynucleotide Kinase according to the manufacturer's instruction (New England BioLabs, Beverly, MA). The recombinant proteins (0.5 to 2 pmol) were mixed with 20 fmol of the radiolabeled probes, without or with variable amounts of unlabeled competitor DNA in reaction buffer (25 mM HEPES-KOH pH7.5, 2.5 mM DTT, 75 mM KCl, 10% glycerol, 1.25 ng poly-dIdC). Each reaction was incubated on room temperature for 10 min without the probes and then incubated on ice for 20 min with the radiolabeled probes. The competitor concentrations were at 0 (-), 0.5 (25X), 1 (50X) and 2 (100X) pmol. After the incubation, the reaction mixtures were resolved by electrophoresis on a 5% non-denaturing polyacrylamide gel. Gels were dried in a gel

dryer (Bio-Rad, Richmond, CA) and exposed to X-ray film (Kodak, Rochester, NY). A list of oligonucleotides for EMSA analysis is provided in Table 3.2.

Maize transgenic plant construction (performed by Dominica Rohozinski and Dr. Frank G. Harmon as a part of the collaborative research)

The construct for *ZmCCA1b* (*ZM2G014902*) overexpression in the binary vector pTF101.1gw1 [92] (generously provided by Dr. Kan Wang, Iowa State University, Ames, Iowa) consisted the *ZmUBI* promoter-*ZmUBI* intron ($ZmUBI_{pro}$) cassette immediately upstream of the *ZmCCA1* coding sequence and the *octopine synthase* (*OCS*) terminator immediately downstream. The *ZmUBIpro* cassette was PCR amplified from the pANIC6A vector [93] with primers F 5'-CACCGTAAGCTTAATGCAGTGCAGCGTGACCC-3' and R 5'-GTAAGCTTTGCAGAAGTAACACCAAACAACAGGGTG-3', which added the *Hin*dIII restriction site (underlined) onto each side of the fragment. The OCS terminator was PCR amplified from the pANIC6A vector with primers F 5'-CCTGCTTTAATGAGATATGCGAGACGC-3' and R 5'-CACCAAAACGACGGCCAGTGCCAA G-3'. The *ZmCCA1b* coding sequence was PCR amplified with primers F 5'-CACCATGGAGGTGAATTCCTCTGGTGAGGAAAC-3' and R5'-TTATGTTGATGCTTCACTATCAAGACGAATCCTCTT-3'. The *ZmUBIpro* cassette was subcloned into the *Hin*dIII site upstream of attR1 site in pTF101.1gw1. The *ZmCCA1b* coding sequence was moved into with LR Clonase II according to the manufacturer's recommendations (Thermo-Fisher, Waltham, Massachusetts). The *OCS*

terminator was subcloned into the *Asc*I site downstream of the *ZmCCA1b* coding sequence. The Plant Transformation Facility (Iowa State University, Ames, Iowa) transformed the complete *ZmCCA1b* overexpression binary construct into the B104 maize inbred by Agrobacterium-mediated transformation according to their published protocol [94]. Pollen from T0 plants, which were confirmed to carry at least copy of the one construct, was used for crosses to the B104 inbred. Transgenic lines were maintained in a hemizygous state by crosses to the B104 inbred.

Plant Genotyping (performed by Dominica Rohozinski and Dr. Frank G. Harmon as a part of the collaborative research)

Transgenic plants were screened using fluorescent probe-based endpoint qPCR. Genomic DNA was extracted from leaf tissue taken from the newest expanded leaf when plants were between V3 and V8. The presence of ADH and BAR was assessed together in PCR reactions with primers 5'-TGTTGAGCAGATCTCGGTGAC-3'and 5'-GTTTCTGGCAGCTGGACTTC-3', with probe 5'-[HEX]AGGACCGGACGGGGCGGTA[BHQ1]-3', for the *bar* gene in pTF101.1gw1; and primers 5'-GAATGTGTGTTGGGTTTGCAT-3' and 5'-TCCAGCAATCCTTGCACCTT-3', with probe 5'-[FAM]TGCAGCCTAACCATGCGCAGGGTA[BHQ1]-3', for ADH1, which is a single copy gene in the maize genome [95]. Samples with PCR amplification for both ADH and BAR were scored as transgenic and those with ADH alone were scored as non-transgenic.

Field and Greenhouse Conditions (performed by Dominica Rohozinski and Dr. Frank G. Harmon as a part of the collaborative research)

Field Trials took place at Oxford Tract in Berkeley, California during summer 2015. The field was sown in a randomized complete block design with two trials separated by one week, planted in late May. In the greenhouse, plants were grown under 16L:8D cycle, supplemented by LumiGrow Pro 325 LED lights (LumiGrow, Inc., Novato, Californa), with 25 °C days and 20 °C nights. Two individual trials were replicated in greenhouse conditions. 5 plants of each family along with 5 non-transgenic siblings were maintained for each trial.

Plant measurements (performed by Dominica Rohozinski and Dr. Frank G. Harmon as a part of the collaborative research)

In field trials, weekly height measurements were taken beginning at the V8 stage until final height was reached. In greenhouse trials, weekly height measurements were taken starting when plants were at the V7 stage until final height was reached. Height was measured from prop roots to the collar of the last fully expanded leaf. Node length was measured at the end of the field or greenhouse trial after plants had reached maturity. Measurements were taken from the upper prop roots to the final visible node, or the base of the tassel.

Chlorophyll measurement

Plant tissue for chlorophyll analysis was harvested using a Harris Uni-Core™ 2.5 mm biopsy punch (Ted Pella, Redding, California). 4 punches were taken midway from the base to the tip of the leaf, equidistant from the mid-vein and the leaf edge. Tissue was

immediately placed in 1 mL DMSO and incubated at 65 °C for 30 min. Absorbance at 645 nm and 663 nm was used to determine chlorophyll content based on Arnon's equation [96].

Phylogenetic analysis

CCA1 homolog proteins used in the analysis are AtCCA1 (At2g46830), AtLHY (At1g01060), SbCCA1 (Sb07g003870), OsCCA1 (Os08g0157600), PnLHY1 (AB429410), PnLHY2 (AB429411), BraA.LHY.a (Bra030496), McCCA1 (AY371287) and GmLCL1 (Glyma11g15580). At: *Arabidopsis thaliana*, Sb: *Sorghum bicolor*, Os: *Oryza sativa*, Pn: *Populus nigra*, BraA: *Brassica rapa*, Mc: *Mesembryanthemum crystallinum*, Gm: *Glycine max*. The protein sequences were aligned using the ClustalW module [97]. The phylogenetic tree was constructed using the Neighbor-Joining method [98]. The bootstrap values were calculated with 1,000 replicates and shown next to the branches. The evolutionary distances were computed using the Poisson correction method [99] and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 472 positions in the final dataset. These procedures were conducted in MEGA6 [100].

Generation of *Arabidopsis* transgenic lines

To create *CAB2*:*LUC*, a native *Arabidopsis CAB2* (*AT1G29920*) promoter fragment was amplified from Col-0 wild-type genomic DNA by the primer pair 5'-CTCGAGTTATATTAATGTTTCGATCATC-3' (*Xho*I) and 5'-CCATGGGTTCGATAGTGTTGGATTATA-3' (*Nco*I) and cloned into pGEM-T (Promega, Madison, WI). After sequence validation, the *CAB2* promoter fragment was

subcloned into pFAMIR binary vector (Basta resistance), which was fused with *LUC* CDS, via *Xho*I/*Nco*I restriction sites. The *CAB2*:*LUC* construct was transformed into the Col-0 and Ws wild-types, respectively, using a floral dip method [101]. Homozygous T3 lines in each background were selected for the following analysis. To generate *CCA1-OX*, *ZmCCA1-OX* and *ZmCCA1b-OX* transgenic lines, *CCA1*, *ZmCCA1a* and *ZmCCA1b* cDNAs were amplified by the following primer pairs, respectively, *CCA1* 5'-CTCGAGATGGAGACAAATTCGTCTGG-3' (*Xho*I) and 5'-GGATCCTCATGTGGAAGCTTGAGTTTC-3' (*Bam*HI), *ZmCCA1a* 5'-CTCGAGATGCCCTTGAGCAATGAG-3' (*Xho*I) and 5'-GGATCCTCATGTTGATGCTTCACTATC-3' (*Bam*HI), and *ZmCCA1b* 5'-CAAGCTCGAGATGGAGGTGAATTCCTCTGGC-3' (*Xho*I) and 5'-GGATCCTTATGTGGATGCTTCGCTATC-3' (*Bam*HI), and cloned into pGEM-T (Promega, Madison, WI). After sequence validation, *CCA1* CDS, *ZmCCA1a* CDS or *ZmCCA1b* CDS was subcloned into pF35SE binary vector (conferring Kanamycin resistance) through *Xho*I/*Bam*HI restriction sites. The resulting constructs designated *CCA1-OX*, *ZmCCA1a-OX* and *ZmCCA1b-OX*, respectively. The construct of vector control, *CCA1-OX*, *ZmCCA1-OX* or *ZmCCA1b-OX* was introduced into the transgenic Col-0 plants that express *CAB2*:*LUC*. T2 lines were used for the luciferase and biomass analysis. For the *cca1-11* complementation analysis with maize *CCA1* homologs, a native *Arabidopsis CCA1* (*AT2G46830.1*) promoter fragment was amplified from Ws wild-type genomic DNA by the primer pair, 5'- GAATTCGCCACGTCCTTCCTTCAATC-3' (*Eco*RI) and 5'- CTCGAGCACTAAGCTCCTCTACACAA-3' (*Xho*I), and cloned into

pGEM-T (Promega, Madison, WI). After sequence validation, the *CCA1* promoter fragment was subcloned into pFAMIR binary vector (Hygromycin resistance), via *Eco*RI/*Xho*I restriction sites. To create *CCA1*:*CCA1*, the *CCA1* CDS was amplified from the Ws cDNA by the primer pair, 5'- ATTTAAATATGGAGACAAATTCGTCTGG-3' (*Swa*I) and 5'- GGATCCTCATGTGGAAGCTTGAGTTTC-3' (*Bam*HI), and cloned into pGEM-T (Promega, Madison, WI). To generate *CCA1*:*ZmCCA1a*, the *ZmCCA1a* CDS was amplified from *Z. mays* B73 cDNA by the primer pair, 5'- CTCGAGATGCCCTTGAGCAATGAG-3' (*Xho*I) and 5'- GGATCCTCATGTTGATGCTTCACTATC-3' (*Bam*HI), and cloned into the pGEM-T (Promega, Madison, WI). To generate *CCA1*:*ZmCCA1b*, the *ZmCCA1b* CDS was amplified from *Z. mays* B73 cDNA by the primer pair, 5'- CAAGCTCGAGATGGAGGTGAATTCCTCTGGC-3' (*Xho*I) and 5- ACCGGATCCTTATGTGGATGCTTCGCTATC-3' (*Bam*HI), and cloned into the pGEM-T (Promega, Madison, WI). After sequence validation, *CCA1* CDS, *ZmCCA1a* CDS or *ZmCCA1b* CDS was subcloned into the pFAMIR (Hygromycin resistance), which harbors the *CCA1* promoter fragment, through the respective restriction sites, generating *CCA1*:*CCA1*, *CCA1*:*ZmCCA1a*, and *CCA1*:*ZmCCA1b* constructs, respectively. The construct of *CCA1*:*CCA1*, *CCA1*:*ZmCCA1a* or *CCA1*:*ZmCCA1b* was transformed into *A. thaliana cca1-11* plants (obtained from the Arabidopsis Biological Resources Center, ABRC, CS9865), which express *CAB2*:*LUC*, generating *CCA1*:*CCA1 cca1-11*, *CCA1*:*ZmCCA1a cca1-11* or *CCA1*:*ZmCCA1b cca1-11* transgenic lines, respectively. T1 transgenic lines were selected for the luciferase analysis.

Arabidopsis plant growth condition and luciferase imaging

Arabidopsis seedlings were plated onto Murashige-Skoog media (Sigma, St. Louis, MO) supplemented with 0.8% agar, 3% sucrose (MS agar) and appropriate antibiotics. After stratification in the dark at 4 °C for 2 days, plates were transferred to a growth room which provided 16L:8D cycles (80 μmol m^2 sec) at 22 °C during 7 days. Survived plants were transplanted to MS medium with no antibiotics for a recovery purpose and grown during 3 days under 16L:8D cycles. The seedlings were transferred to 96-well white microtiter plates (Nunc, Rochester, NY), which contain MS agar, 0.33 mM D-luciferin (Gold Biotechnology, Manchester, RI), and then grown under constant light (80 μmol m^2 sec). The next day, luminescence recordings were performed on a TopCount (Packard Bioscience, Shelton, CT) over 7 days, and the data were analyzed by fast Fourier transform–nonlinear least squares (FFT–NLLS) [102] using the interface provided by the Biological Rhythms Analysis Software System version 3.0 (BRASS) (http://www.amillar.org).

3.3 Results

Maize *CCA1* homologs exhibit diurnal expression patterns and binding activities

Photosynthetic activities and starch metabolism are diurnally regulated in maize [56, 57] as in *Arabidopsis* diploids, hybrids and allopolyploids [31], suggesting a role for the circadian clock in promoting seedling growth in maize, as in *Arabidopsis*. To test this, I investigated *CCA1* homologs and their functions in maize. Being an ancient tetraploid [103, 104], maize has paralogous duplicates of many clock gene homologs [58]. The two

homologs, designated *ZmCCA1a* (*ZM2G474769*) and *ZmCCA1b* (*ZM2G014902*) and located on chromosomes 10 and 4, respectively [58], are closely related to *CCA1* and *LHY* in *Arabidopsis* (Figure 3.1A and B). In the maize gene model (B73 RefGen_v2), *ZmCCA1a* is a truncated version; in the ~20-kb upstream region, there is another homologous gene (*ZM2G175265*) that encodes a MYB DNA-binding domain protein (Figure 3.2). The short transcript (*ZM2G474769*, *ZmCCA1a*) could be a splicing variant and was used for expression and other analyses in this dissertation study. In *Arabidopsis*, CCA1 lacking the MYB DNA-binding domain acts as a dominant negative factor to fine-tune the period length [105]. The two maize genes exhibited similar expression patterns, although *ZmCCA1b* was expressed at higher levels than *ZmCCA1a* [58], in most of 17 maize tissues examined [90] (Figure 3.1C), indicating that *ZmCCA1b* may play a larger role than *ZmCCA1a*. Expression of *ZmCCA1b* and *ZmCCA1a* was diurnally regulated and peaked at ZT3 (Figure 3.2A), whereas *GIGANTEA1* (*gi1*), *ZmPRR59*, and *ZmTOC1a* [58, 61], peaked at ZT12 (Figure 3.3). Interestingly, all clock genes analyzed were nonadditively expressed (deviated from MPV) in the hybrids at time-points that coincided with nonadditive photosynthetic and metabolic activities during the day (Figure 2.3). In the hybrids, both *ZmCCA1a* and *ZmCCA1b* were up-regulated in the hybrids in the morning phase, while *gi1* and *ZmPRR59*, a *PRR5* homolog in *Arabidopsis*, were up-regulated in the hybrids at ZT9. However, *ZmTOC1a* was down-regulated in the hybrids in the middle of the day and at night. These data indicate altered expression of circadian clock genes in the maize hybrids as in *Arabidopsis* allotetraploids [35] and intraspecific hybrids [36, 106].

Altered *CCA1* expression may affect abundance of CCA1 that binds to the promoters of other clock and output genes through evening element (EE) or CCA1-binding site (CBS) [35, 107]. To test DNA-binding activities of ZmCCA1a and ZmCCA1b, I expressed and purified fusion proteins of maltose binding protein (MBP) with either ZmCCA1a or ZmCCA1b from *Escherichia coli*, generating recombinant ZmCCA1a (rZmCCA1a) and ZmCCA1b (rZmCCA1b), respectively. I then performed the electrophoretic mobility shift assay (EMSA) with the purified recombinant proteins. The EMSA showed that ZmCCA1b but not ZmCCA1a (excluding the MYB DNA-binding domain) bound competitively and specifically to the EE and CBS elements of *ZmTOC1a*, *gi1* and *ZmPRR59* promoter fragments *in vitro* (Figure 3.2C). The binding signal depended on EE or CBS elements as well as the recombinant ZmCCA1b (rZmCCA1b) protein concentration. When the EE or CBS was mutated (M) or no EE or CBS (N) was present in the endogenous promoter fragments, the binding activity of rZmCCA1b was undetectable. As a control, maltose-binding protein (MBP) showed no binding activities. Like the control, rZmCCA1a showed no binding activity. The absence of the MYB DNA-binding domain in truncated ZmCCA1a likely explains undetectable target binding. These data suggest that ZmCCA1b is a functional protein that binds to the promoters of circadian clock genes via *cis*-acting EE and CBS elements.

Functional conservation of maize circadian clock genes

To test the function of maize CCA1 homologs, I generated *Arabidopsis CCA1*, maize *ZmCCA1a* or *ZmCCA1b* overexpression (OX) lines in the transgenic *A. thaliana* (Col-0) plants that also express a luciferase (LUC) reporter driven by the promoter of

CHLOROPHYLL A/B BINDING PROTEIN 2 (*CAB2*) (*CAB2*:*LUC*). These lines designated *CCA1-OX*, *ZmCCA1a-OX*, and *ZmCCA1b-OX*, respectively (Figure 3.2D). While the wild-type and transgenic control (*Vec-1*) lines showed rhythmic *CAB2*:*LUC* expression, overexpressing *CCA1* abolished *CAB2*:*LUC* expression rhythmicity under constant light (Figure 3.2D and Figure 3.4A), as previously reported [51]. Interestingly, in the *ZmCCA1b-OX* lines the *CAB2*:*LUC* expression rhythmicity was dampened, and the timing of peak expression delayed under the constant light. Overexpressing *ZmCCA1a* in the *ZmCCA1a-OX* lines had a smaller effect on *CAB2*:*LUC* expression rhythmicity, and the expression peak was lower than that in the *ZmCCA1b-OX* lines (Figure 3.2D and 3.3A). The weaker *ZmCCA1a-OX* phenotype may be associated with the lack of binding activity in the truncated ZmCCA1a to the EE or CBS of gene promoters (Figure 3.2C)

As a result of disrupting the clock gene functions, aerial biomass of *CCA1-OX* and *ZmCCA1b-OX* lines was significantly lower than the wild-type or the transgenic control (Figure 3.2E), consistent with the growth disadvantage in clock gene mutants and overexpression lines [50, 51]. Biomass in the *ZmCCA1a-OX* lines was not significantly different from the controls, suggesting that a full-length clock gene is required for growth vigor.

Expressing *ZmCCA1a* or *ZmCCA1b* under the *Arabidopsis CCA1* promoter partially rescued the early period phenotype of *CAB2*:*LUC* in the *cca1-11* mutant (Figure 3.4B and C). Under constant light, the period of *CAB2*:*LUC* rhythms was shorter in the *cca1-11* mutant (23.9 ± 0.1 h) than in the wild-type (25.4 ± 0.3 h). Expressing *Arabidopsis CCA1* in the *cca1-11* mutant, as a control, fully rescued the short period

phenotype of *CAB2*:*LUC* (25.4 ± 0.1 h). Expressing *CCA1*:*ZmCCA1b* (24.8 ± 0.1 h) or *CCA1*:*ZmCCA1a* (24.2 ± 0.1 h) in the *cca1* mutant also lengthened the period relative to *cca1-11* (23.9 ± 0.1 h). These data suggest that the circadian clock function between *Arabidopsis* and maize *CCA1* homologs is conserved. *ZmCCA1b* complemented the mutant phenotype, while *ZmCCA1a* did not, suggesting that DNA binding activities conferred by the MYB DNA-binding domain are required for the clock function.

Overexpressing *ZmCCA1b* reduces growth vigor in maize

To test if *ZmCCA1b* regulates growth vigor in maize, *ZmCCA1b* driven by the constitutive *UBIQUITIN* promoter-intron cassette (*UBI*) [108] was expressed in maize B104 line. Multiple independent transgenic lines were generated; two lines, designated *OX1-3* and *OX10-1*, were analyzed in both greenhouse and field conditions (Figure 3.5). These lines were maintained as heterozygotes to prevent transgene silencing and were subsequently genotyped to identify the lines carrying the transgene for analysis. In the greenhouse, plant height was dramatically reduced in the OX lines (Figure 3.5A). The *OX1-3* lines exhibited a more severe phenotype than the *OX10-1* line (Figure 3.5B). Reduction in plant height was associated with reduced lengths of early nodes (from #3 to #10), while the total node number was unaffected (Figure 3.5C). The *OX1-3* line showed higher expression levels of *ZmCCA1b* transgene and more severe reduction in plant height than the *OX1-3* line (Figure 3.5D). Moreover, the amount of total chlorophyll and chlorophyll a and b was reduced in the *OX1-3* line in both greenhouse and field conditions, with more severe reduction in the greenhouse than in the field. In the field, the *OX1-3* line also showed reduced height (Figure 3.5F), but the reduction in node

length occurred in the later nodes (#7 to #16) (compare Figure 3.5G with 3.5C). Moreover, the phenotype was less severe in the field than in the greenhouse. A major difference between the greenhouse and field conditions is the light intensity, suggesting a role of light compensation for circadian regulation in growth vigor, which remains to be investigated. Together, these data suggest that disrupting circadian regulation reduces chlorophyll content and growth vigor in maize as in *Arabidopsis* (Figure 3.2E) [35, 50]

3.4 Discussion

CCA1, along with LHY, is considered as a central clock regulator with important roles in plant growth and physiology [42, 43, 45]. In this chapter, I firstly, to my knowledge, characterize molecular and circadian functions of maize *CCA1* homologs. These data imply a functional conservation of *CCA1* homologs between the *Arabidopsis* and maize homologs, and further demonstrate that the MYB DNA-binding domain in ZmCCA1s is a key for the normal clock function. According to these results, I propose that the circadian clock mediated by *CCA1* is a conserved mechanism in coordinating plant growth and physiological responses to the daily changes across the plant species and, more importantly, that the nonadditive expression of *ZmCCA1s* in the maize hybrids could affect growth vigor.

Clock regulator homologs in plant species, including rice [109-111], poplar [112], brassica [113], soybean [114], sorghum [60], barley [115, 116] and maize [58, 117], display robust rhythmic gene expression patterns, which are similar with their *Arabidopsis* counterparts, suggesting functional conservation of the clock regulators

across these plant species despite highly divergent phylogenetic relationships [118]. The results with *ZmCCA1a* and *ZmCCA1b* extend the simple implication to detailed molecular characterizations of the clock function, of which the center is the DNA-binding activities conferred by the MYB DNA-binding domain. For example, the MYB DNA-binding domain deficient ZmCCA1a failed not only to bind to promoters of other clock regulator genes, bur also to perform the normal clock function *in planta* unlike CCA1 and ZmCCA1b, of which both contain the domain. This indicates that ZmCCA1b, and possibly the full-length ZmCCA1a, is a functional clock regulator as its *Arabidopsis* homolog. It is reasonable to consider that the nonadditive expression of *ZmCCA1a* and *ZmCCA1b* in the hybrids causes different levels, or patterns, of the clock activity, which coordinates important biological activities including growth and fitness. Then, the remained question could be how ZmCCA1s, as central clock regulators, affect expression of output genes and growth vigor in maize hybrids. In this regard, it will be of special interest to reveal the architecture of ZmCCA1-regulatory networks for heterosis in maize, in order to understand the molecular connection of the circadian clock with growth heterosis in maize and its biological significance.

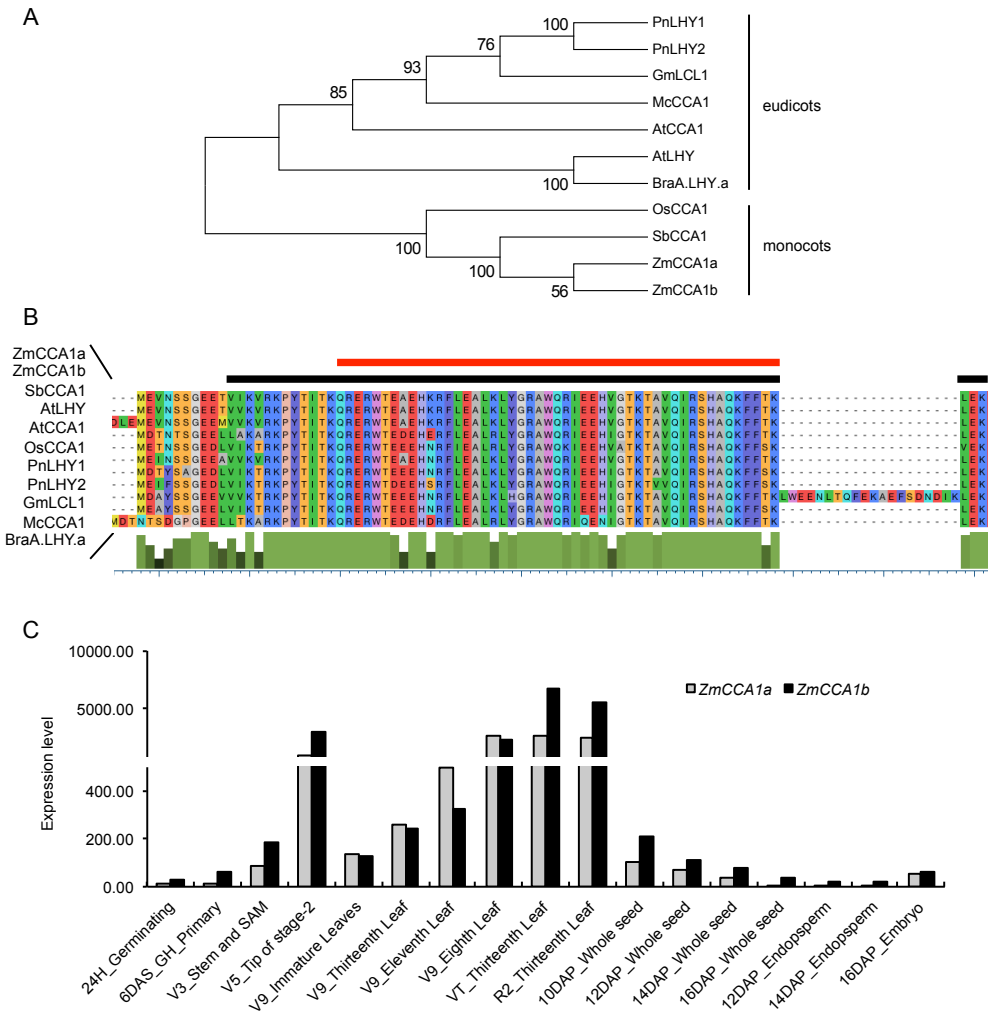

Figure 3.1 Phylogenetic tree and multi-alignment of CCA1 homologs in plants, and gene expression patterns in maize tissues.

(A) Phylogenetic tree of CCA1/LHY homologs in plants. The Neighbor-Joining phylogenetic tree of CCA1/LHY was constructed from amino acid sequences, and bootstrap values calculated with 1,000 replicates are shown next to the branches. (B) Multiple sequence alignment showing N-terminus MYB DNA-binding domain of CCA1 homologs in plants. Amino acid sequences were aligned using the ClustalW module. The MYB DNA-binding domain, indicated by the black bar, is highly conserved in the CCA1 homologs. The red bar indicates the region recognized by αCCA1. Consensus match is plotted below. Representatives are shown from monocots (Os, *O. sativa*; Sb, *S. bicolor*; Zm, *Z. mays*) and eudicots (At, *A. thaliana*; Bra, *B. rapa*; Mc, *M. crystallinum*; Pn, *P. nigra*; Gm, *Glycine max*). (C) Expression levels of the two maize *CCA1* paralogous genes in 17 tissues of B73. The maize gene atlas developed by RNA-seq was used to compare the tissue-specific expression for *ZmCCA1a* and *ZmCCA1b*. Expression level indicates fragments per kilobase pair of exon model per million fragments mapped (FPKM).

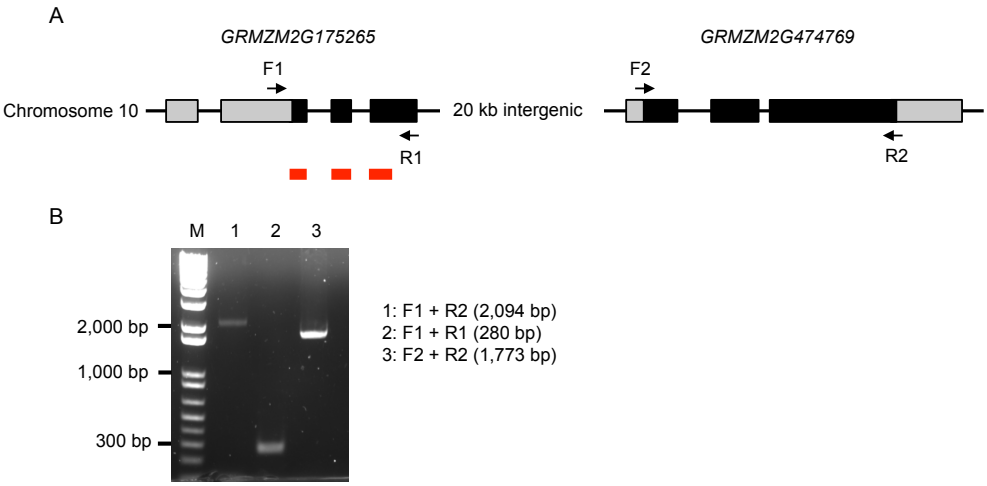

Figure 3.2 PCR amplifications of *ZmCCA1a* transcripts.

(A) Gene structure of *ZmCCA1a*, which is a combination of *GRMZM2G175265* and *GRMZM2G474769*. Note that the two genes annotated in the maize gene model (B73 RefGen_v2) are separated by a 20 kb intergenic region. Black and grey boxes indicate coding and UTR regions, respectively. Red boxes indicate the coding region that encodes the highly conserved N-terminus MYB DNA-binding domain as shown in Figure 3.1B. (B) Semi-quantitative RT-PCR of co-transcripts of *GRMZM2G175265* and *GRMZM2G474769* (lane 1), transcripts of *GRMZM2G175265* (lane 2) and transcripts of *GRMZM2G474769* (lane 3). M indicates a DNA ladder. Primer pairs and expected size of fragment PCR-amplified are shown in the right. Location of each primer is shown in (A).

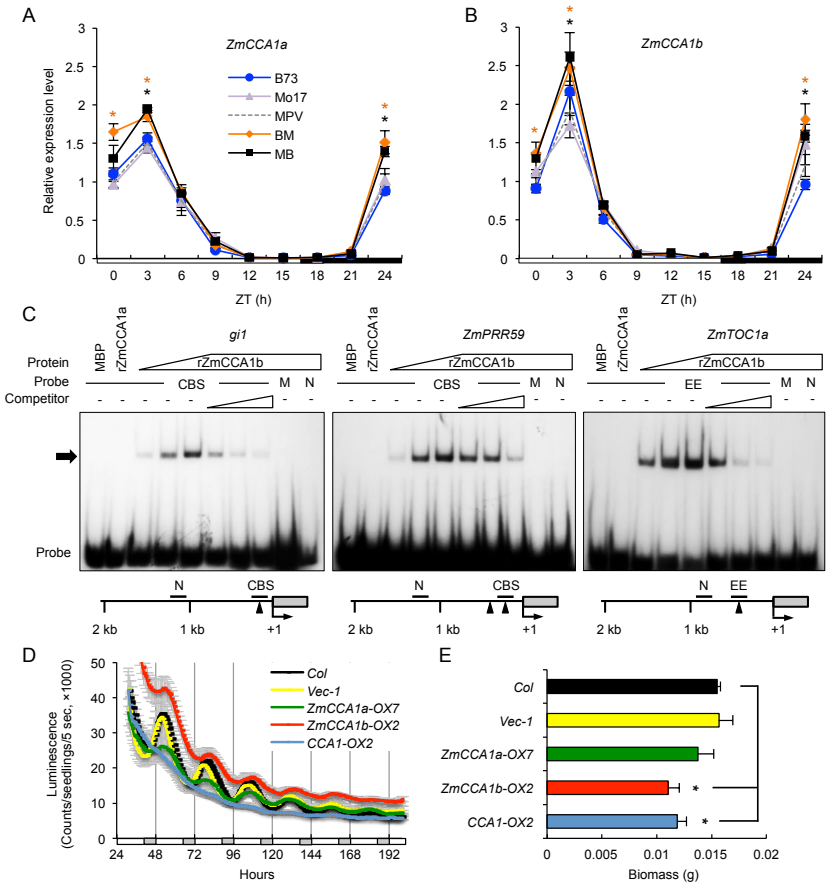

Figure 3.3 Molecular characterizations of maize *CCA1* homologs.

(A and B) Relative expression levels (means ± SEM, n = 3) of *ZmCCA1a* (A) and *ZmCCA1b* (B) every 3 hours in a 24-hour period (light/dark cycle is shown below the histogram). Significant difference between hybrids and MPV was calculated using Student's t-test, *p-value < 0.05. The relative expression level in MPV at ZT0 was set to 1. (C) Binding of recombinant ZmCCA1b (rZmCCA1b) to promoters of putative maize clock genes *in vitro*. Radioisotope-labeled DNA probes (endogenous promoter fragments) were incubated in the presence of MBP (1 pmol), rZmCCA1a (1 pmol), and rZmCCA1b (0.5, 1 and 2 pmol). Shifted protein-DNA complexes are indicated by the arrow. Competitors: 25X, 50X and 100X molar excess of unlabeled prompter DNA. M: DNA in which EE or CBS site was mutated; N: no EE or CBS site in the DNA fragment. Location of probes for each gene is shown below the gel image. Arrowheads represent EE or CBS site. Numbers are relative to the transcription start site (+1) and 5' UTR (grey box). (D) *CAB2*:*LUC* activity rhythms in wild-type (Col-0), *Vec-1*, *ZmCCA1a-OX7*, *ZmCCA1b-OX2* and *CCA1-OX2* under constant light (means ± SEM, n = 12-16). White and grey bars represent the subjective day and night, respectively. T2 plants were used in the analysis. (E) Dry aerial biomass of wild-type (Col-0), *Vec-1*, *ZmCCA1a-OX7*, *ZmCCA1b-OX2* and *CCA1-OX2*. Dry biomass (grams) was measured 35 DAP in the diurnal conditions (means ± SEM, n = 10).

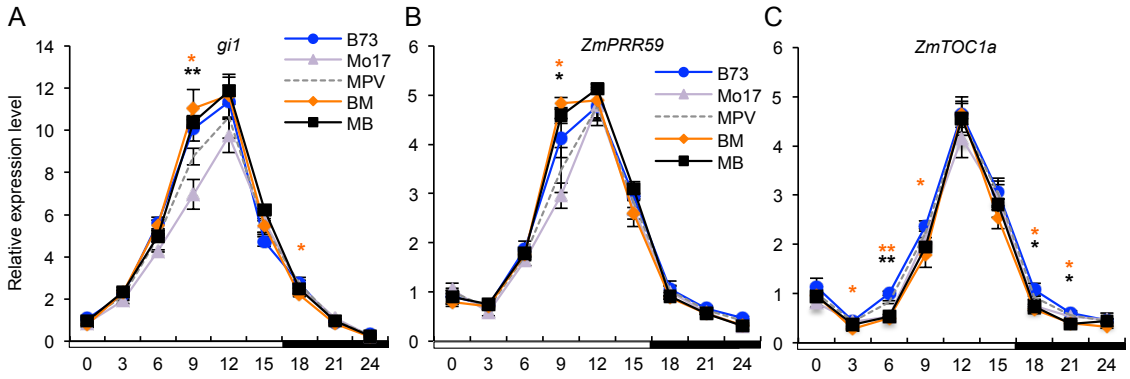

Figure 3.4 Diurnal expression levels of putative clock genes in the maize inbreds and hybrids.

Relative expression levels (means ± SEM, n = 3) of *gi1* (A), *ZmPRR59* (B) and *ZmTOC1a* (C) every 3 hours in a 24-hour period (light/dark cycle is shown below the histogram). Significant difference between hybrids and MPV was calculated using Student's t-test, *p-value < 0.05. The relative expression level in MPV at ZT0 was set to 1.

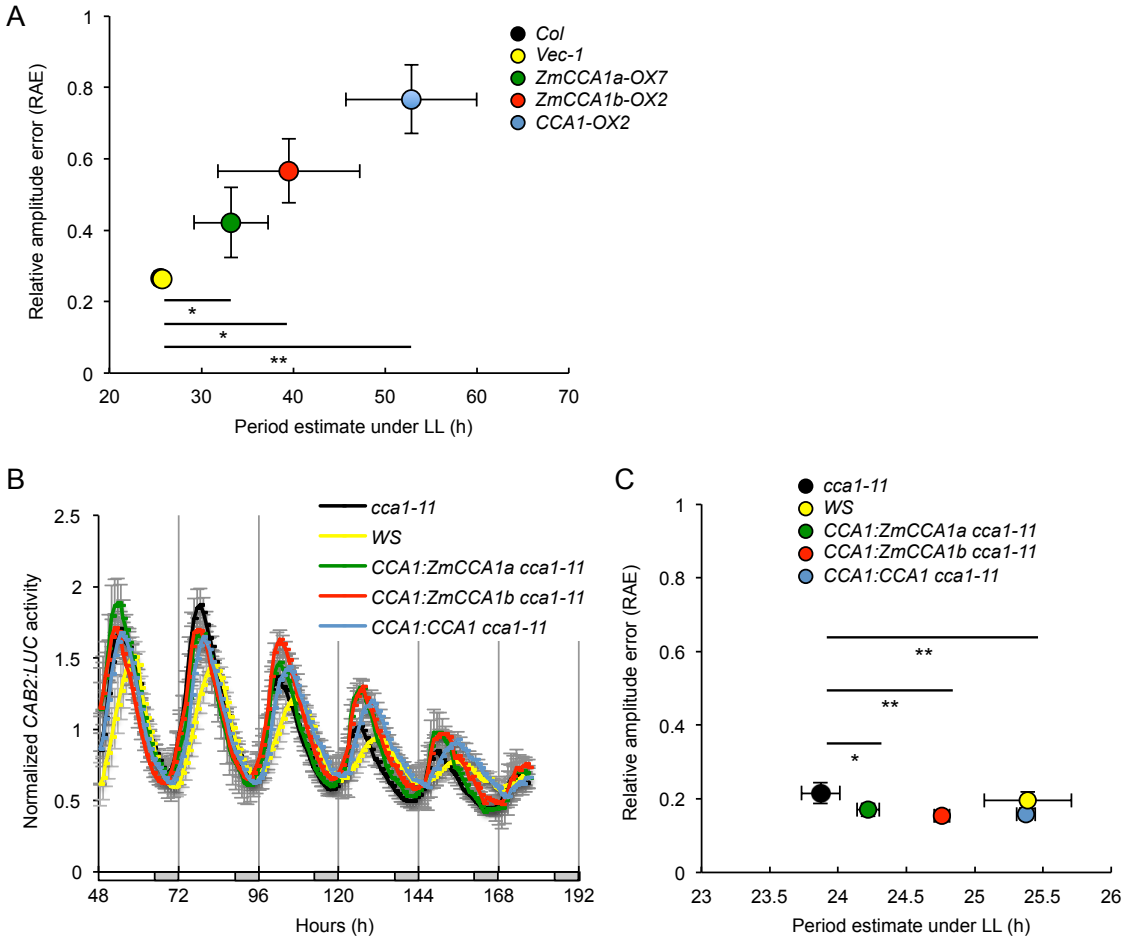

Figure 3.5 Circadian characterization of maize *CCA1* homologs in *Arabidopsis*.

(A) A plot of period versus relative amplitude error (RAE) of CAB2:LUC activity in wild-type (Col-0), *Vec-1*, *ZmCCA1a-OX7*, *ZmCCA1b-OX2* and *CCA1-OX2* under constant light (LL) (means ± SEM). RAE > 0.6 represents arrthymicity. Mean values were compared to those of the vector control line using Student's t-test. Single and double asterisks indicate significant differences at p-value < 0.05 and p-value < 0.01, respectively. T2 plants were used in the analysis. (B) *CAB2*:*LUC* activity rhythms in wild-type (WS), *cca1-11*, *CCA1*:*ZmCCA1a cca1-11*, *CCA1*:*ZmCCA1b cca1-11* and *CCA1*:*CCA1 cca1-11* under LL (means ± SEM, n = 4-8). White and grey bars represent the subjective day and night, respectively. T1 plants were used in the analysis. T1 plants were used in the analysis. (C) A plot of period versus relative amplitude error (RAE) of *CAB2*:*LUC* activity in wild-type (WS), *cca1-11*, *CCA1*:*ZmCCA1a cca1-11*, *CCA1*:*ZmCCA1b cca1-11* and *CCA1*:*CCA1 cca1-11* under LL (means ± SEM). Mean values were compared to those of *cca1-11* using Student's t-test, *p-value < 0.05 and **p-value < 0.01.

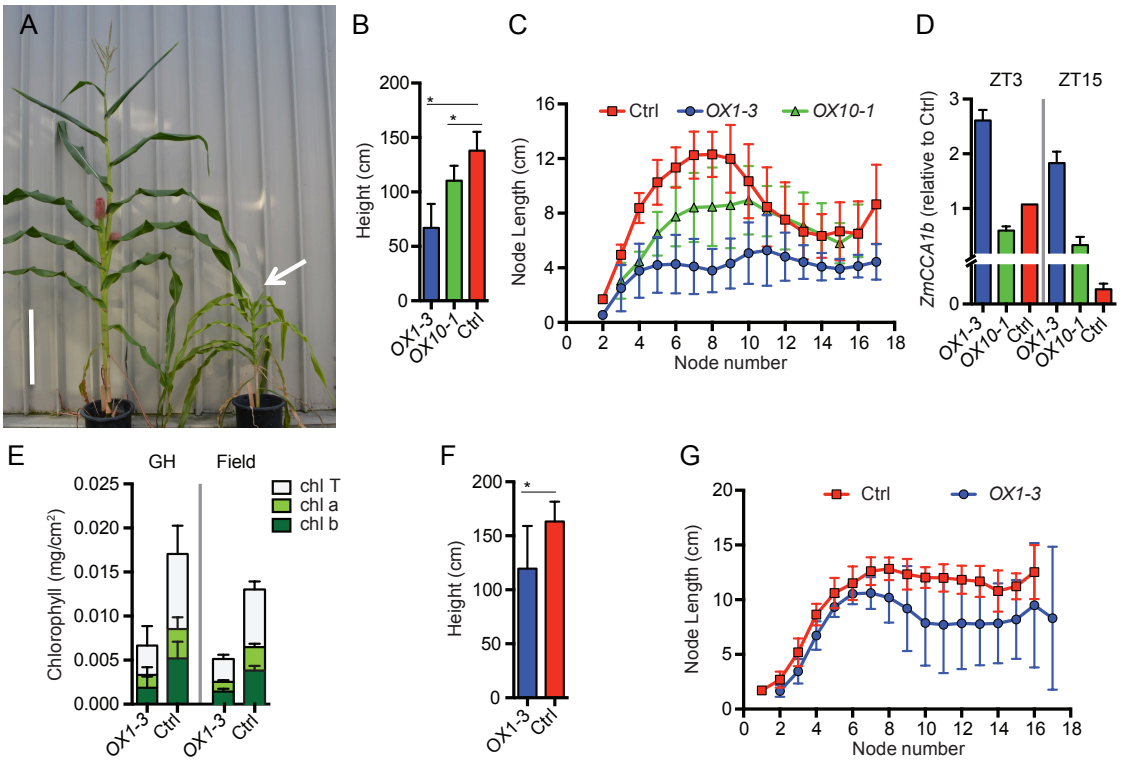

Figure 3.6 Maize transgenic plants that overexpressed *ZmCCA1b*.

(A) Phenotypes of *OX1-3* and control lines grown in the greenhouse (an arrow indicates the reduced height in *OX1-3*). Scale bar = 40 cm. (B) Relative expression levels (to the control, Ctrl, at ZT3) of *ZmCCA1b* in the leaves of *OX1-3* and *OX1-10* transgenic plants at ZT3 and ZT15 (n = 2, ± SD). Significant difference was calculated using Student's t-test, *p-value < 0.001. (C) Height at maturity for *OX1-3*, *OX10-1*, and control plants grown in the field. (D) Measurement of node elongation for *OX1-3, OX10-1*, and control plants grown in the field (n = 10, ± SD). (E) Chlorophyll a (chl a), chlorophyll b (chl b), and total chlorophyll (chl T) content in leaves of *OX1-3* and control plants grown in the greenhouse (GH) or field as indicted. Chlorophyll measurements for greenhouse grown and field-grown plants should not be compared since different developmental stages were tested between the two growth conditions. Greenhouse-grown plants were at stage V9 and field-grown plants were at maturity. (F) Height at maturity for *OX1-3* and control plants grown in a greenhouse. Significant difference was calculated using Student's t-test, *p-value < 0.001. (G) Measurement of node elongation for *OX1-3* and control plants grown in the greenhouse (n = 10, ± SD).

This figure was generated by Dominica Rohozinski and Dr. Frank G. Harmon as a part of the collaborative research.

Name	Gene ID	Direction	Sequence
F1	*GRMZM2G175265*	Forward	CTTTTTGCTGGATTTGGAG
F2	*GRMZM2G474769*	Forward	GAATTCATGCCCTTGAGCAATGAG
R1	*GRMZM2G175265*	Reverse	CTAGACTGAACCAAACAAGAG
R2	*GRMZM2G474769*	Reverse	GTCGACTCATGTTGATGCTTCACTATCAAGACG

Table 3.1 List of primers used in semi-quantitative RT-PCR (5' to 3')

Name	Gene ID	Forward sequence	Reverse sequence
ZmCCA1a	GRMZM2G474769	TTGGTGAGCCAAGGGCTTCCTTT	CTGTTCGGCGCAATTTCAGCTT
ZmCCA1b	*GRMZM2G014902*	CGAAGCATCCACATAATTGATTT	GCACTGCATTGCAAGATCTGA
gi1	*GRMZM2G107101*	GCAATACCAGCTCATTGGATAGTGT	CAACATCGCCATTCAGTAGGACCTC
ZmPRR59	*GRMZM2G135446*	TCTTGCTGCCTCCCAATGACCATA	ACTACTTGCACCAGCATCTTCCCT
ZmTOC1a	*GRMZM2G020081*	GCCAACCAATACGGATGTCAT	TGGATCGTCATCTTCGTCTTCA
18srRNA	*AF168884*	TCTGTGATGCCCTTAGATGTTCTG	CTGTCGGCCAAGGCTATATACT

Table 3.2 List of primers used in qRT-PCR (5' to 3')

Name	Gene ID	Description	Sequence
gi1	*GRMZM2G107101*	F	GCATCGCGGAAAAAAATCTTTTTCTCGCC
		R	GGCGAGAAAAAGATTTTTTCCGCGATGC
		M_F	GCATCGCGGAAAAAAgagTTTTCTCGCC
		M_R	GGCGAGAAAActcTTTTTTCCGCGATGC
		N_F	GTCCGAGGCTAGGCTCGGGTGAAGCGTG
		N_R	CACGCTTCACCCGAGCCTAGCCTCGGAC
ZmPRR59	*GRMZM2G135446*	F	CTCTTTAAAAAAAAATCTTCATTTAAGG
		R	CCTTAAATGAAGATTTTTTTTTAAAGAG
		M_F	CTCTTTAAAAAAAAAgagTCATTTAAGG
		M_R	CCTTAAATGActcTTTTTTTTTAAAGAG
		N_F	GGGTCAGGCCGTCACGGGTCCAATGGAC
		N_R	GTCCATTGGACCCGTGACGGCCTGACCC
ZmTOC1a	*GRMZM2G020081*	F	GGCGAGTTTTAGATATTTTCACCCGCTCC
		R	GGAGCGGGTGAAAATATCTAAAACTCGCC
		M_F	GGCGAGTTTTccccgcgaTCACCCGCTCC
		M_R	GGAGCGGGTGAtcgcggggAAAACTCGCC
		N_F	GACCCCACCTAGCCGGTCTGTGTCTGTCC
		N_R	GGACAGACACAGACCGGCTAGGTGGGGTC

Table 3.3 List of oligonucleotides used in EMSA (5' to 3')

EE or CBS elements are underlined. Mutated sequences are lower cases. F: forward strand; R: reverse strand; M: DNA in which EE or CBS site was mutated; N: no EE or CBS site in the DNA fragment.

Chapter 4. Genome-wide binding profiles of ZmCCA1s in maize inbred lines and F1 hybrids

4.1 Introduction

As transcriptional regulation by transcription factors (TF) is considered as the primary level of gene regulation [119], understanding how and when the core clock regulators, as transcriptional factors, coordinately regulate rhythmic transcription of clock output genes has been of great interest in the study of clock biology. In recent years, chromatin immunoprecipitation followed either by microarray (ChIP-chip) or deep sequencing (ChIP-seq) has emerged as a standard technique for profiling protein DNA-binding analysis *in vivo* [120-122]. Particularly, increasing access to and decreasing costs of next-generation sequencing technology in the past decade have allowed ChIP-seq to be the most preferred approach for TF-target binding profiling. Despite its technical and computational challenges [123-125], ChIP-seq is already valuable due to its versatility; it allows us not only to obtain binding targets, but also to identify DNA binding motifs and obtain quantitative readouts for comparative analysis (e.g. comparing the number of peaks or enrichments across genotypes or time-points) [121, 126].

In *Arabidopsis*, ChIP-seq analyses of circadian clock regulators, including CCA1, TOC1, PRR5 and PRR7, have made key contributions to the current understanding of the circadian clock, as in mammals [127-129]. For example, recent ChIP-seq studies of CCA1 demonstrated that CCA1 bound to more than hundreds of genes associated with diverse biological pathways under circadian and diurnal conditions [130, 131]. The ChIP-

seq analysis of TOC1 revealed hundreds of circadian target genes mostly expressed during the morning phase, including *CCA1* and *LHY*, suggesting that TOC1 acts as a general repressor of clock-output gene expression not as the previously suggested activator [132]. The ChIP-seq analysis of PRR5, combined with genome-wide gene expression profiles, demonstrated that it coordinates expression timing of output genes, in which transcription factors are highly enriched, associated with diverse physiological processes [133]. However, whether maize utilizes the circadian clock in the similar way remains unknown.

In Chapter 2 and Chapter 3, I identified that ZmCCA1b, not the truncated ZmCCA1a, is a functional central clock regulator and further that the MYB DNA-binding domain is a key for clock functions in growth vigor. I proposed that the altered expression of *ZmCCA1a* and *ZmCCA1b* in the maize hybrids is associated with growth heterosis. In this regard, identifying and comparing its gene-regulatory networks in the maize hybrids and inbreds are desirable to understand how ZmCCA1s affect expression of their output genes and ultimately growth heterosis.

Here, I perform genome-wide diurnal binding profiles of ZmCCA1s in the maize inbreds and hybrids using ChIP-seq analysis. I report a temporal shift of ZmCCA1-binding activities toward early morning in the hybrids relative to the inbreds, suggesting that the activation of morning-phased genes in the hybrids promotes photosynthesis and growth vigor. This temporal shift in activation leads to nonadditive gene expression that establishes heterosis, followed by maintenance of heterosis by additive gene expression.

These data provides novel insights, suggesting a molecular mechanism that reprograms circadian-regulatory networks to promote growth vigor in maize hybrids.

4.2 Materials and methods

Chromatin immunoprecipitation (ChIP) and ChIP-qPCR

For more detailed information for this method, see Chapter 5. Aerial tissues of three biological replicates were harvested at ZT3, ZT9 and ZT15 and were completely submerged in fresh prepared formaldehyde buffer (0.4 M Sucrose, 10 mM Tris-HCl, 1 mM PMSF, 3% Formaldehyde, 5 mM β-mercaptoethanol). A vacuum was applied for 20 min; after adding glycine to a final concentration 125 mM, another vacuum was applied for 5 min. After the formaldehyde/glycine buffer was removed, the cross-linked tissues were washed with sterilized water, briefly dried with paper towels, immediately frozen in liquid nitrogen and stored at -80 °C for experimental use. Sonication was performed with 10 cycles of 30 s pulses on and 30 s pulses off to achieve an average fragment size of 400-bp using a Bioruptor Sonicator (Diagenode, Sparta, NJ). Samples were subjected to centrifugation at 13,800 *g* at 4 °C for 10 min, and the supernatant containing chromatin was transferred to a new 1.5 ml tube. Immunoprecipitation (IP) was performed using 600 µl of sonicated chromatin with 5 µg of anti-CCA1 antibody. For each IP sample, a mock (no antibody) and input (no IP) were included. Purified DNA (IP, mock and input) from ChIP was resuspended in TE, pH 7.5 and used for qPCR and ChIP-seq library preparation.

ChIP-seq libraries for ChIP and input samples from two biological replicates were constructed using the standard NEB protocol (New England BioLabs, Beverly, MA) using custom made adapters containing barcodes used to pool multiple samples for sequencing. ChIP DNA was subject to end repair, dA-tailing, ligation with the adapters, and amplification by 18 cycles of PCR using Next High-Fidelity 2XPCR Master Mix (New England BioLabs, Beverly, MA). Pair-end (2X100-bp) sequencing was performed on an Illumina HiSeq 2500 at The University of Texas at Austin Genomic Sequencing and Analysis Facility.

For qPCR, purified ChIP and mock DNA was diluted 2 times, and input DNA was diluted 5 times. The diluted DNA (2 µl) was used for qPCR in an ABI7500 machine (Applied Biosystems, Foster City, CA) using FastStart Universal SYBR Green Master (ROX) (Roche Applied Science, Indianapolis, IN). Enrichment of the binding in IP and mock samples was normalized relative to the corresponding input sample. Primers are shown in Table 4.3.

ChIP-seq data analysis

Raw reads were subjective to quality trimming and adaptor clipping using FASTX-Toolkit (hannonlab.cshl.edu/fastx_toolkit) followed by removing orphan reads. The filtered pair-end reads were mapped to the maize reference genome (Zmays_284_AGPv2, release 5b.60) and Mo17 genome (reference-guided assembly based on the B73 genome) [103, 134], using Bowtie (version 2.1.0) [135], allowing 1 mismatch in the 20-bp 'seed' with options '--score-min L,0,-0.3 -X 1000 --no-mixed --no-discordant'. Particularly, for B73 and Mo17, the filtered pair-end reads were mapped

to the maize B73 reference genome and the Mo17 genome, respectively. For F1 hybrids, the filtered pair-end reads were mapped to both B73 and Mo17 reference genomes, and the best alignment was selected for each read. Only reads concordantly mapping to the genome exactly 1 time was kept and used for peak calling. Duplicate reads were removed from the Bowtie output using samtools rmdup command [136]. To normalize sequencing depth of the mapped reads among samples from different genotypes and time-points, we adopted a normalization method based on the previously published paper [137]. In order to adjust different sequencing depths among genotype and time-points, the uniquely and concordantly mapped paired-end reads were "down-sampled" to the lowest number of samples without PCR duplicates (Appendix). For example, if B73, Mo17, BM and MB had 4, 5, 6 and 7 million reads respectively, all samples were down-sampled to 4 million reads. This will ensure that numbers of binding peaks are comparable among genotypes and time-points. The criteria for identifying targets were based on abundance of peak enrichment using ANOVA test (p-value < 0.05 as a cut-off). When target genes showed same abundance 2 or more time-points, they were excluded from the phase-shift analysis.

The latest version of model-based analysis of ChIP-seq algorithm (MACS, version 2.0.1) [138] was used to identify enriched peaks on each ChIP-seq file against the corresponding input file using a mappable genome size of –g 2.07e+09 and cut off p-value of 1e-3. Peaks were defined if they overlapped in two biological replicates. To make a master-peak list from the three time-points, the peaks obtained from each time-point were merged for each genotype. Integrated Genomics Viewer (version 2.3.46) [139] was used to visualize duplicate-filtered input subtracted ChIP signals. For the

normalization, modules of the deepTools suite (http://deeptools.ie-freiburg.mpg.de) [140] were used. *De novo* motif analysis was performed with the total master-peaks using MEME suite [141] in which JASPAR CORE (2014) plants database was selected for motif database. GO enrichment analysis was performed using the web interface of agriGO (http://bioinfo.cau.edu.cn/agriGO/) [142] with false-discovery rate adjusted p-value < 0.05 (Hypergeometric test) as a cutoff. Biological process among ontology categories was used. The heatmap for GO analysis (Figure 4.4C) was generated using R package gplots. Peak distribution was determined with respect to the maize gene model (release 5b.60).

Gene expression analysis

For gene expression analysis, aerial tissues were harvested and immediately frozen in the liquid nitrogen. Total RNA was extracted from the frozen tissues by PureLink^{TM} Plant RNA reagent (Invitrogen, Carlsbad, CA) and treated with RQ1 DNaseI (Promega, Madison, WI) according to the manufacturers' instructions. For cDNA synthesis, 1 µg of DNaseI-treated total RNA was incubated with Omniscript reverse transcriptase (Qiagen, Valencia, CA) in the presence of 10 µM random hexamer (GeneLink, Hawthorne, NY). For qPCR, FastStart Universal SYBR Green Master (ROX) (Roche Applied Science, Indianapolis, IN) was used in the presence of gene-specific primers and template cDNAs in an ABI7500 (Applied Biosystems, Foster City, CA) or LightCycler 96 machine (Roche Applied Science, Indianapolis, IN). The control was *18S rRNA* to estimate the relative expression levels of each gene in three biological replicates. A list of primers for gene expression analysis is provided in Table 4.1.

Purification of recombinant ZmCCA1b and ZmCCA1a proteins

The coding sequence (CDS) of *ZM2G474769* (*ZmCCA1a*) was amplified from B73 cDNA by the primer pair 5'-GAATTCATGCCCTTGAGCAATGAG-3' (*Eco*RI, underlined) and 5'-GTCGACTCATGTTGATGCTTCACTAT-3' (*Sal*I). The full-length *ZmCCA1b* (*ZM2G014902*) CDS was amplified from B73 cDNA by the primer pair 5'-GGATCCATGGAGGTGAATTCCTCTGGC-3' (*Bam*HI) and 5'-GTCGACTTATGTGGATGCTTCGCTATC-3' (*Sal*I). The *ZmCCA1a* or *ZmCCA1b* cDNA fragment was cloned into a pGEM-T (Promega, Madison, Wisconsin). After sequence verification, the *ZmCCA1a* or *ZmCCA1b* CDSs was subcloned into pMAL-C2 (New England BioLabs, Beverly, MA) through *Eco*RI/*Sal*I and *BamH*I/*Sal*I restriction sites, respectively. *E. coli* strain Rosetta-gami B competent cells (Novagen, Madison, WI) was used to transform empty pMAL (expressing maltose-binding protein, MBP), pMAL-ZmCCA1a or pMAL-ZmCCA1b, which were grown in 4 ml of Luria-Bertani (LB) media with Carbenicillin (100 mg/L) at 37 °C for 18 h. The overnight cultures of Rosetta-gami B cells containing pMAL, pMAL-ZmCCA1a or pMAL-ZmCCA1b construct were diluted into 1:100 in 80 ml LB media with Carbinicillin (100 mg/L) and grown at 37 °C to an OD_{600} value of 0.5, when isopropyl-β-D-thiogalactoside (IPTG) (0.1 mM) was added. After 20 h of additional incubation at 16 °C, cells were harvested after centrifugation at 4,000 *g* at 4 °C for 10 min and resuspended in 2 ml of Column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA). After frozen at -20 °C for 18 h, cells were lysed by a Bioruptor Sonicator (Diagenode, Sparta, NJ) and centrifuged at 20,000 *g* at 4 °C for 20 min. The cleared cell lysates were diluted 1:5 with Column buffer, and

loaded on amylose-coupled agarose resin columns prepared according to the manufacturer's instruction (New England BioLabs, Beverly, MA). After columns were washed with 12 volumes of Column buffer, MBP, rZmCCA1a and rZmCCA1b were eluted with elution buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM Maltose). After filtration by Amicon Ultra 100 K (Millipore, Darmstadt, Germany), the purified MBP, rZmCCA1a or rZmCCA1b was aliquoted and stored at -80 °C.

Electrophoretic mobility shift assays (EMSA)

DNA probes were generated by annealing PAGE-purified sense and antisense oligonucleotides (Table 4.2). The double-stranded oligonucleotides were ^{32}P end-labeled using a T4 Polynucleotide Kinase according to the manufacturer's instruction (New England BioLabs, Beverly, MA). The recombinant proteins (0.5 to 2 pmol) were mixed with 20 fmol of the radiolabeled probes, without or with variable amounts of unlabeled competitor DNA in reaction buffer (25 mM HEPES-KOH pH7.5, 2.5 mM DTT, 75 mM KCl, 10% glycerol, 1.25 ng poly-dIdC). Each reaction was incubated on room temperature for 10 min without the probes and then incubated on ice for 20 min with the radiolabeled probes. The competitor concentrations were at 0 (-), 0.5 (25X), 1 (50X) and 2 (100X) pmol. After the incubation, the reaction mixtures were resolved by electrophoresis on a 5% non-denaturing polyacrylamide gel. Gels were dried in a gel dryer (Bio-Rad, Richmond, CA) and exposed to X-ray film (Kodak, Rochester, NY). A list of oligonucleotides for EMSA analysis is provided in Table 4.2.

4.3 Results

Maize ZmCCA1s bind to EE and CBS variants and Dof-binding motifs

These data indicate that ZmCCA1s, like their homologs in *Arabidopsis*, act as central clock regulators in maize. We predicted that altered *ZmCCA1* expression in the F1 hybrids plays a role in maize heterosis. To test how ZmCCA1s affect expression of output genes and growth vigor in maize, we raised antibodies against the N-terminus MYB DNA-binding domain (Figure 3.1B), which could recognize ZmCCA1b and a full-length ZmCCA1a if the MYB DNA-binding domain is included. Western blot analysis indicated that accumulation of ZmCCA1s reached high levels in the morning, peaking at ZT3, and decreased during the day (Figure 4.1A). Using the antibodies, we tested genome-wide ZmCCA1-binding profiles in the F1 hybrids and their parents at ZT3, ZT9 and ZT15 using chromatin immunoprecipitation (ChIP) followed by deep sequencing (ChIP-seq). Sequencing reads were mapped onto the B73 reference genome (AGPv2) and the Mo17 genome, respectively [103] (see Materials and methods). The paired-end reads that were uniquely and concordantly mapped were normalized among replicates and all genotypes using a "down-sampled" approach [138] (Figure 4.1B and Appendix). The peaks that were present in both replicates (Pearson correlation coefficient = 0.97, based on read coverage over consecutive 1-kb bins) in one or more genotypes were used for further analysis. The number of peaks ranged from 1,874 to 3,364 with a total of 10,136 peaks in all four genotypes (Figure 4.2A). Compared to total genomic features, the peaks were enriched in coding, upstream including 2-kb promoter and 5' untranslated region (UTRs), and downstream flanking sequences (Figure 4.2B). *De novo* motif analysis identified two top-scoring motifs, namely, AAAATA, an EE (AAATATCT) variant, and

AAGAAA, a CBS (AAAAATCT) variant, which represented 81% and 65% of total binding peaks, respectively (Figure 4.3A). The sequence variants may suggest divergence of canonical EE and CBS sequences between *Arabidopsis* (eudicot) and maize (monocot) over 150-200 million years of evolution [143]. Further analysis using the TOMTOM tool in MEME [144] found a Dof-binding motif (AAAGC), which was statistically significantly similar to the non-classified motif (p-value = 0.019) (Figure 4.3A). The Dof-transcription factor genes are involved in tissue-specific expression, light signaling, and carbon fixation in maize [145, 146], suggesting that ZmCCA1 interacts with other transcription factors to mediate light-regulated gene expression.

Temporal regulation of ZmCCA1-binding activities in maize F1 hybrids and inbreds

These peaks were associated with target genes within 10-kb sequences using the filtered gene set (FGS) in MaizeGDB (http://archive.maizegdb.org/cgi-bin/termrefs.cgi?id=2366450). The number of ZmCCA1-target genes ranged from 1,406 to 2,511 in each genotype, resulting a non-redundant set of 4,319 target genes (Figure 4.5), which is consistent with the observation that 10-20% of maize genes exhibit diurnal expression [54, 58], although not all diurnally expressed genes are regulated by ZmCCA1. ZmCCA1s bound to EE and CBS elements of putative circadian clock homologs (Figure 4.2B), as shown in the EMSA results (Figure 3.2C). These putative clock genes were nonadditively expressed at certain time points in the hybrids relative to the parents (Figure 3.3). These data suggest a role for ZmCCA1-binding activities in circadian-mediated gene expression. In the reciprocal crosses, more ZmCCA1-binding

peaks were found in BM than in MB (Figure 4.4A), which may suggest a parent-of-origin effect of circadian clock function, consistent with that in *Arabidopsis* intraspecific hybrids [36].

Interestingly, the proportion of ZmCCA1-binding peaks in the hybrids was significantly shifted towards ZT3 (Fisher's exact test, p-value < 4.3E-15 at ZT3; p-value < 4.2E-05 at ZT9), while the proportion in the inbreds was relatively unchanged between ZT3 and ZT9 (Figure 4.3B). The ZmCCA1-binding target genes, associated with ZmCCA1-binding peaks, were partitioned into shared among all genotypes and specific to either hybrids or inbreds and among all time-points (Figure 4.5A) or at each time-point (Figure 4.5B). Among the shared set of ZmCCA1-binding target genes (418) (significant overlaps; Fisher's exact test, p-value < 1.59E-179), the proportion was significantly more at ZT3 in the hybrids than in the inbred lines (Fisher's exact test, p-value < 7.2E-12) but fewer at ZT9 in the hybrids than in the inbred lines (Fisher's exact test, p-value < 0.03) (Figure 4.3C). The temporal shift of ZmCCA1-binding target genes in the inbred lines were not statistically significant (Figure 4.3C), but the shift was significant for the targets that were not shared among all genotypes (Figure 4.4A). Similarly, among the ZmCCA1-binding target genes that were specific to hybrids or inbreds, the temporal shift was significantly in the hybrids but not in the inbreds (Figure 4.4B). These data suggest that the ZmCCA1-binding activities of target genes in the hybrids have shifted towards the early morning. However, this does not exclude a possibility of many targets that are preferentially bound by ZmCCA1s in the hybrids compared to the inbreds.

ZmCCA1-binding genes in Gene Ontology (GO) groups of energy and metabolism

To test the biological function of the temporal shift, we classified ZmCCA1-binding target genes into GO groups (Figure 4.5). The shared target genes in both hybrids and inbreds were enriched in the genes in electron transport chain, generation of precursor metabolites and energy, photosynthesis, and light reaction among all time-points (Figure 4.5A) or in different time-points (Figure 4.5B). This indicates a major role of ZmCCA1s in modulating photosynthetic and metabolic activities, which is consistent with the circadian control of energy and metabolism in other plants [31, 35] and mammals [147]. Moreover, the hybrid-specific target genes were significantly enriched in the genes involved in protein and cellular protein metabolic processes, indicating a role for ZmCCA1s in altered cellular metabolism that would promote growth vigor in maize hybrids. Notably, the inbred-specific target genes among all time points or in different time points were overrepresented in the genes involved in intracellular transport and protein localization (Figure 4.5), which could suggest a role for protein stability and movement in maintaining cellular growth and development [148].

Notably, the enrichment of carbon fixation genes in F1 hybrids was almost exclusive at ZT3 (Figure 4.3D), whereas these genes in the inbreds also occurred in other time-points, in addition to ZT3. The genes in GO terms of other pathways, including protein catabolic process, tRNA metabolic process and protein transport, were also enriched, but they were not associated with the phase-shift. This suggests a uniform shift in activating photosynthetic and carbon metabolism pathways to the early morning in the hybrids.

Expression of carbon fixation genes is consistent with the phase-shift of ZmCCA1-binding activities in the hybrids

Morning phased-genes associated with the temporal shift included those encoding a starch synthase III (*ZM2G121612*), a light harvesting complex photosystem II (*ZM2G033885*), a malate dehydrogenase (*ZM2G129513*), and a starch synthase II (*ZM2G126988*), a putative phospholipid-transporting ATPase (*ZM2G398288*) (Figure 4.6A-C and G and Figure 4.7A). These genes were up-regulated in the morning and in the hybrids compared to the inbreds, which correlated with the shift of ZmCCA1-binding activities in the hybrids (Figure 4.6D-F and Figure 4.8E). Thus, the phase-shift of ZmCCA1-binding in the hybrids is positively associated with expression levels of these morning-phased genes. EMSA assays have confirmed the binding activities of ZmCCA1b to the promoters of *ZM2G121612*, *ZM2G033885* and *ZM2G129513* (Figure 4.8).

For the afternoon-phased genes, including *gi1* and *ZM2G412611* encoding an alpha-glucan water dikinase chloroplast precursor, their expression peaks were also shifted from ZT12 to ZT9 (Figure 4.6G-I), which correlated with the temporal shift of ZmCCA1-binding in the hybrids (Figure 4.7F). Collectively, these results indicate that altered temporal binding activities of ZmCCA1s to the clock output genes in the maize hybrids are responsible for the expression rhythms of carbon fixation genes, promoting photosynthesis and metabolism.

There were a few exceptions. For example, the expression shift for the genes encoding a photosynthetic reaction center protein (*ZM2G427369*) and a photosystem II

reaction center protein Z (*ZM2G394732*) was consistent with the ZmCCA1-binding shift in the hybrids (Figure 4.7B and C), but they were expressed at higher levels in most time points examined. The transcript level of the gene (*ZM2G448142*) encoding an ATP synthase b-subunit in the hybrids was similar to that in the inbreds (Figure 4.7D). These data suggest that other circadian clock regulators such as TOC1 and ELF3 homologs may also play a role in the diurnal regulation of gene expression in maize hybrids.

Validation of the temporal shift of ZmCCA1-binding targets in the hybrids

The ChIP-seq binding activities of ZmCCA1s to the promoters of several carbon fixation genes were confirmed by ChIP-qPCR (Figure 4.6D-F) and EMSA (Figure 4.8) assays. To test if the temporal shift of ZmCCA1-binding targets is induced in the hybrids, we examined ZmCCA1-binding activities to the promoters of selected four carbon fixation genes in B104, *OX1-3* and *OX1-3*XMo17 by ChIP-qPCR (Figure 4.9). Compared to the control (B104), overexpressing *ZmCCA1b* in the *OX1-3* increased the binding levels at the four genes tested. Remarkably, in the F1 hybrid (*OX1-3*XM), the ZmCCA1-binding activities were not only increased in levels but also shifted to the morning-phase (ZT3), which was not obvious in the *OX1-3* line. Although not all possible genotypes were tested, the data suggest a temporal sift of ZmCCA1-binding activities in the hybrids towards early morning, which could play an important role in maize heterosis. Collectively, these results indicate that altered temporal ZmCCA1-binding activities to the clock output genes in the maize hybrids are responsible for expression rhythmicity of carbon fixation genes, promoting photosynthesis, metabolism and growth vigor.

A developmental role in heterosis

Circadian regulation of gene expression in *Arabidopsis* intraspecific hybrids is established during embryo development and maintained during seedling growth [36], and biomass heterosis is established during early stages of seedling development [80, 81]. Consistent with this notion, expression of the circadian clock genes (*ZmCCA1a* and *ZmCCA1b*) and their output genes in maize hybrids is developmentally regulated during seedling stages (Figure 4.7G). Expression of *ZmCCA1a*, *ZmCCA1b* and carbon fixation genes was nonadditive in maize hybrids at 5 DAP and 8 DAP but gradually became additive at 11 DAP and 14 DAP. As a control, *Cell Number Regulator 2* (*CNR2*), which controls cell numbers in maize [149], was nonadditively expressed in the hybrids late inseedling development (14 DAP). These data suggest that the influence of circadian regulation on seedling heterosis in maize is likely developmentally regulated, corresponding to gene expression dynamics during maize development [146].

4.4 Discussion

Developmental regulation and a temporal shift model for heterosis

Heterosis is predicted to arise from allelic interactions between parental genomes, leading to altered regulatory networks that promote the growth and fitness of hybrids [3, 31, 32]. Here, I demonstrate that one such regulator is the circadian clock in maize hybrids. The circadian clock genes are functionally conserved in *Arabidopsis* and maize. In hybrids, the maize central clock proteins target thousands of output genes early in the morning, including carbon fixation genes. The data collectively support a phase-shift model for heterosis (Figure 4.10). ZmCCA1-binding targets the regulatory networks

involved in energy and metabolism, which is established early in the seedling and subsequently maintained during growth. During establishment, morning-phased bindings of ZmCCA1s to carbon fixation genes in F1 hybrids (relative to the inbreds) may cause nonadditive gene expression; consequently nonadditive increases in carbon fixation rate, and, ultimately, biomass accumulation. During late stages of development, gene expression in the hybrids is shifted towards additive expression of the circadian-mediated carbon fixation and metabolic genes. This shift of different gene expression modes may explain different findings of additive expression [150], nonadditive, and/or all modes of gene expression [151], which have been documented in maize hybrids or hybrids with different ploidy levels [152]. The temporal shift from non-additive to additive gene expression has also been observed in *A. thaliana* F1 hybrids [153, 154]. Moreover, ZmCCA1-binding target genes are dependent on genotypes (inbreds vs. hybrids) (Figure 4.5). In the hybrids, coordinated regulation of the protein and cellular metabolic genes could allow them to achieve greater protein metabolic efficiency, as many cellular metabolic processes are simultaneously engaged in production of stable or efficient metabolites, saving energy used in their synthesis and metabolism [155]. Our findings provide novel insights into a clock-based mechanism for a temporal shift of morning-expressed genes in the maize hybrids to promote photosynthetic and metabolic activities, leading to biomass heterosis.

Circadian regulation and biomass heterosis in maize

Circadian clocks regulate biological processes of most organisms including plants and animals [156]. Plant growth and development, from stress responses, biomass

accumulation, to seedling growth, flowering, and fruit formation, is directly controlled by clock regulators through transcriptional and post-transcriptional regulation of output processes [45, 157]. Physiological activities, including CO_2 fixation and starch metabolism, are diurnally regulated by circadian clock factors. CCA1 serves as both a transcriptional repressor and activator of target genes, including those in the core circadian oscillator and output pathways [157, 158]. For example, CCA1 is a repressor for evening-phased genes in *A. thaliana*; it directly binds to promoters of the evening-phased central clock genes (*TOC1* and *GI*) and represses their expression [159, 160]. In the *cca1-11* mutant, expression of evening-phased starch metabolic genes is upregulated [161]. Meanwhile, CCA1 is an activator for morning-phased genes; it directly binds to the promoters of the morning-phased central clock genes (*PRR7* and *PRR9*) [162] and cold stress responsive genes [163], which are repressed in *cca1-11* and *cca1 lhy* mutants. CCA1 also regulates expression of morning-phased photosynthetic genes, including *CAB* genes, and their expression is suppressed in the *cca1* mutant [164].

Our study has demonstrated functional conservation between *Arabidopsis* and maize CCA1 homologs. Consistent with the positive role of CCA1 in morning-phased genes in *Arabidopsis*, ZmCCA1-bindings to the morning-phased carbon fixation genes were positively correlated with the expression of these genes in the morning (ZT3), whereas ZmCCA1 association with promoters of these genes in the afternoon (ZT9) were negatively correlated with the expression of these genes. It is likely that the positive and negative roles of CCA1 are achieved through its interaction with other factors. For example, in *A. thaliana*, CCA1 interacts with LHY *in vitro* and *in vivo*, and the two

factors function synergistically in circadian clock regulation [165]. A splice variant of CCA1 (CCA1β), which lacks the MYB DNA-binding domain, interferes with the transcriptional function of CCA1 by interacting with the full-length CCA1 (CCA1α) or LHY to form nonfunctional heterodimers [166]. Without *CCA1β*, the circadian period is longer, whereas overexpressing *CCA1β* leads to a short-period phenotype, suggesting that interacting with spliced CCA1β fine-tunes CCA1α function. In maize, although the long-form of *ZmCCA1a* is not studied, the short-form of *ZmCCA1a* lacking the MYB DNA-domain could act as a dominant negative factor to fine-tune circadian rhythms in maize as in *Arabidopsis* [166], as the effect of this splicing variant in *Arabidopsis* overexpressing and complementation tests is relatively small compared to *ZmCCA1b* (Figure 3.2 D and E and Figure 3.4). Both maize *ZmCCA1b* and *ZmCCA1a* are up-regulated in the hybrids in a morning-specific manner with higher expression of *ZmCCA1b*, which may serve as co-regulators for each other with redundant and/or separate functions in the transcriptional regulation of output (photosynthetic, carbon fixation, and metabolic) genes in the hybrids.

It is notable that the F1 hybrids share some similar molecular and physiological phenotypes such as photosynthetic rate and expression of carbon-fixation genes with one of the parent B73. However, the hybrid still performs better than the best parent, which is known as best parent-heterosis (BPH), instead of mid-parent heterosis (MPH) [32]. This suggests that some molecular events could be more important than others in promoting growth vigor. For example, net carbon gain (starch and sugar contents) in the sink and source correlates better with the level of heterosis than other molecular and physiological parameters (Figure 2.3). Interestingly, the genes corresponding to starch metabolic

pathways are regulated by the circadian clock and nonadditively expressed, which could play an important role in heterosis. Further testing the functions of these genes using transgenic and genetic (mutant) approaches will reveal mechanisms for the circadian clock to regulate unique biological pathways that stimulate growth vigor in maize hybrids.

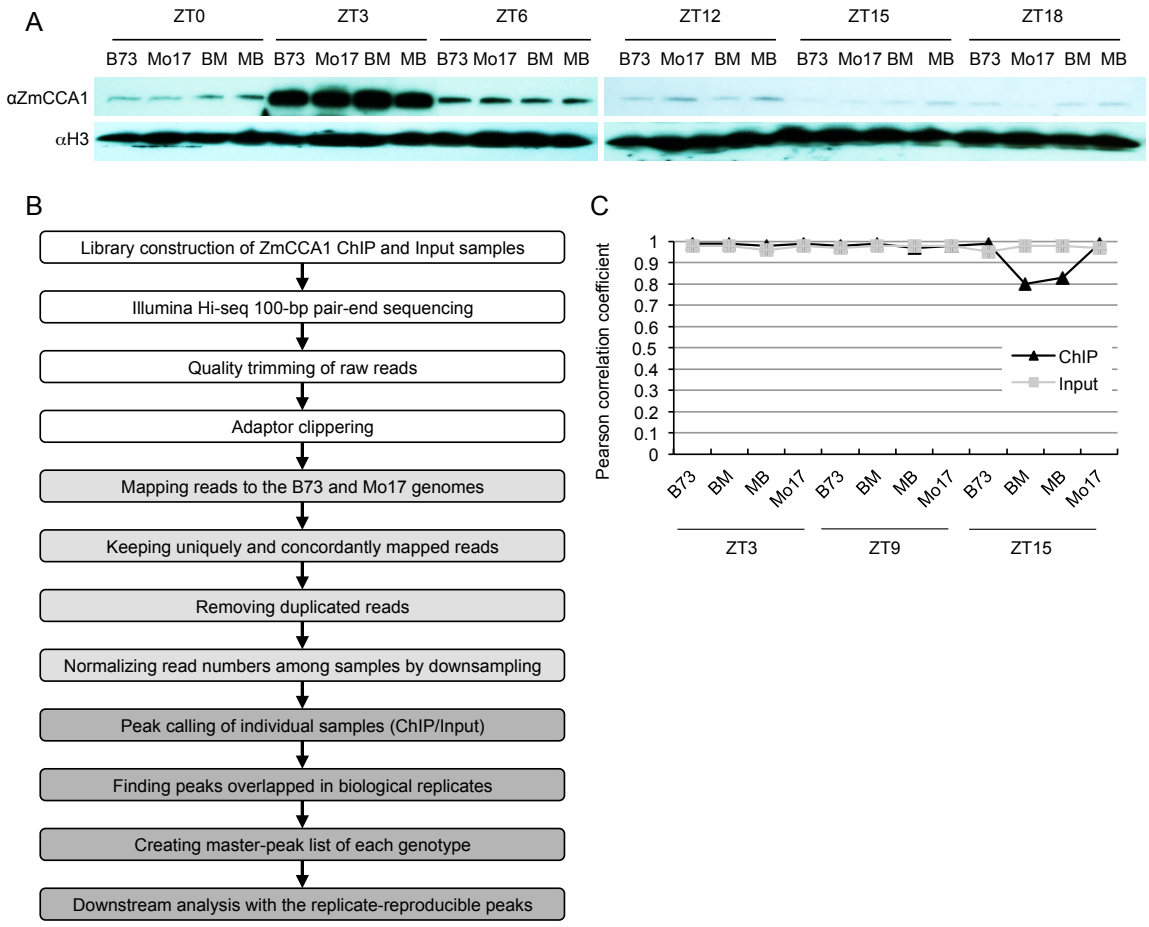

Figure 4.1 Specificity of antibody against ZmCCA1s and computational pipeline for ChIP-seq analysis.

(A) Immunodetection of ZmCCA1s in the maize inbreds and hybrids using the polyclonal antibodies used for our ChIP-seq. Histone H3 protein was used as a loading control. Plant tissues were collected at time points as labeled under the diurnal condition and used for protein extraction. (B) Workflow of computational analysis of ZmCCA1 ChIP-seq. Summary of mapping statistics is provided in Appendix. (C) Line graph displays pairwise correlations between biological replicates calculated by Pearson correlation coefficients in windows of 1-kb over mapped read files.

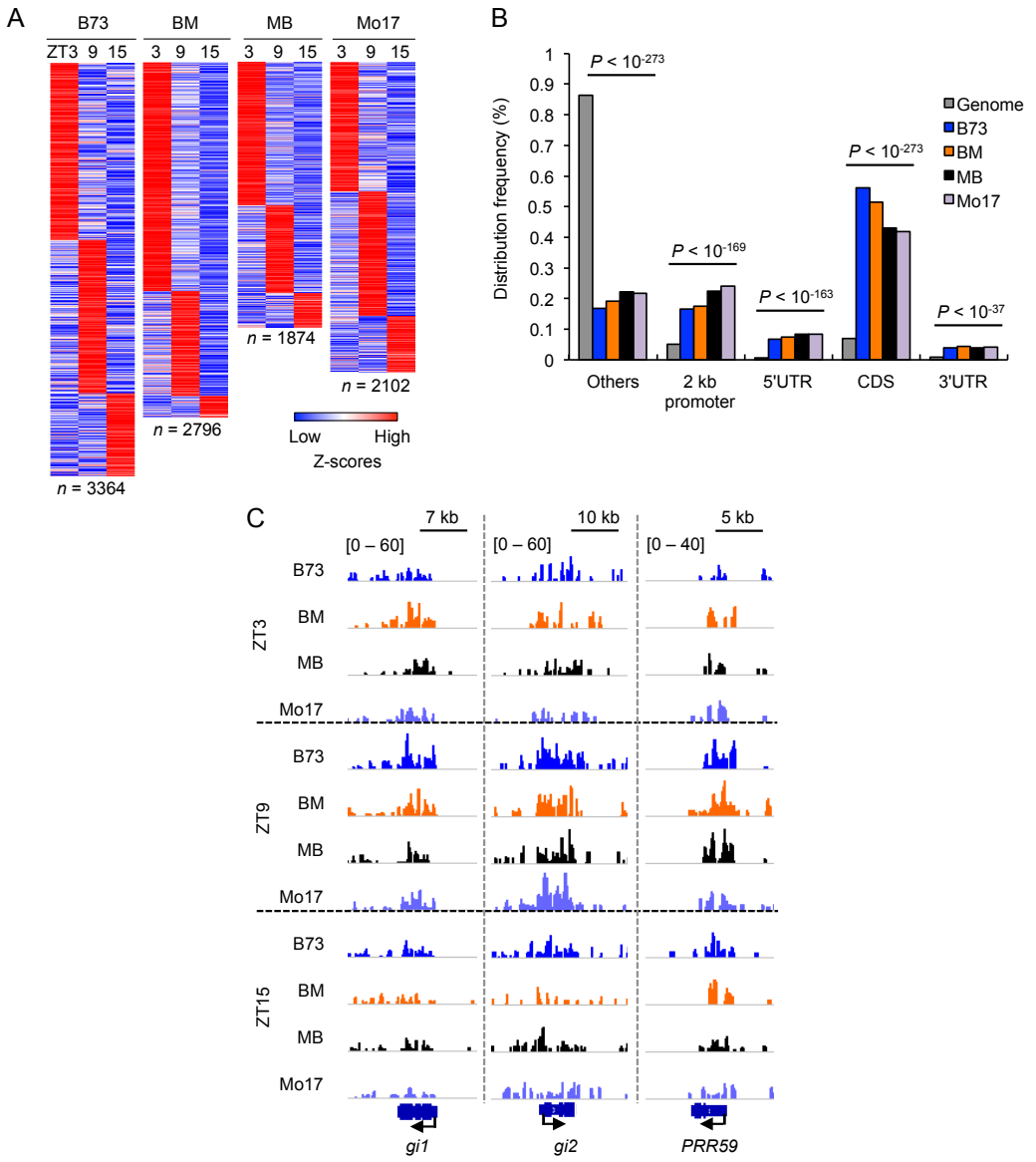

Figure 4.2 Relative distribution of ZmCCA1-binding peaks across genomic regions and binding motifs found in the peaks.

(A) Hierarchically clustered heatmap showing the ChIP signals of ZmCCA1-binding (Z-scores) on ZmCCA1-binding peaks of each genotype. (B) Histogram showing genomic annotation of ZmCCA1-binding peaks. Genome indicates the maize genome fraction as a control. Significant differences between Genome and each genotype are shown above each category using Fisher's exact test. (B) Local enrichment of ZmCCA1-binding at putative maize clock genes at ZT3, ZT9 and ZT15. The Y-axis indicates input-subtracted read density on a same-scale for all genotypes and time-points. Arrows indicate gene orientation.

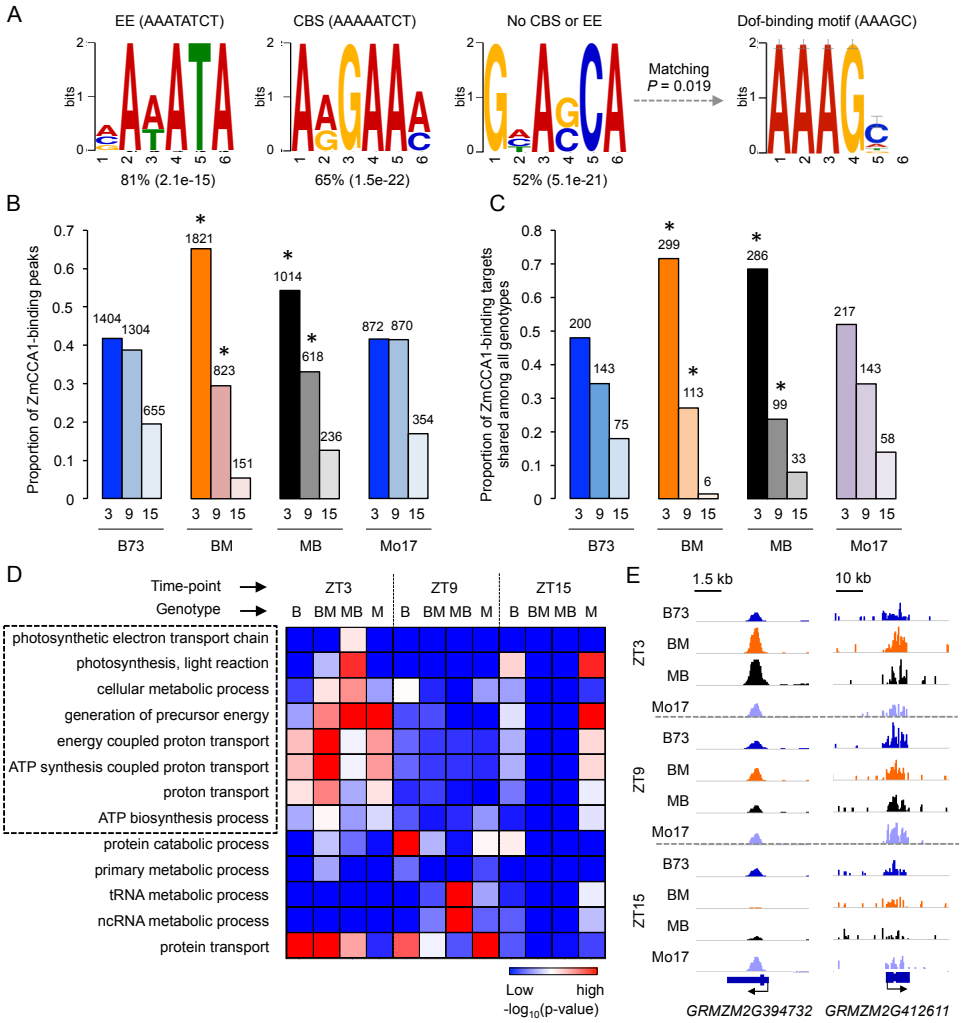

Figure 4.3 ChIP-seq analysis of ZmCCA1s in the maize inbreds and F1 hybrids.

(A) *De novo* motif analysis of ZmCCA1-binding sites using MEME suites showed enrichment of EE, CBS, and no CBS or EE motifs. In the latter, Dof-binding motif was significantly matched in TOMTOM analysis. Percentage (p-value) is shown below each motif. (B and C) Proportion of ZmCCA1-binding peaks (B) and binding targets (C) shared in all genotypes at each time-point. Asterisks indicate statistical significance of ZT3 or ZT9 frequencies between the F1 hybrids and their parents were calculated by Fisher's exact test. The number of peaks or genes is shown above each bar. (D) Gene Ontology (GO) classification of ZmCCA1-binding targets in each genotype at each time point by GO analysis (false-discovery rate adjusted p-value < 0.05, Hypergeometric test). GO terms associated with carbon fixation are marked by a dashed box. (E) Examples of altered temporal bindings of ZmCCA1s in F1 hybrids compared to the inbred lines. The Y-axis indicates input-subtracted read density on a same-scale for all genotypes and time-points. Arrows indicate gene orientation.

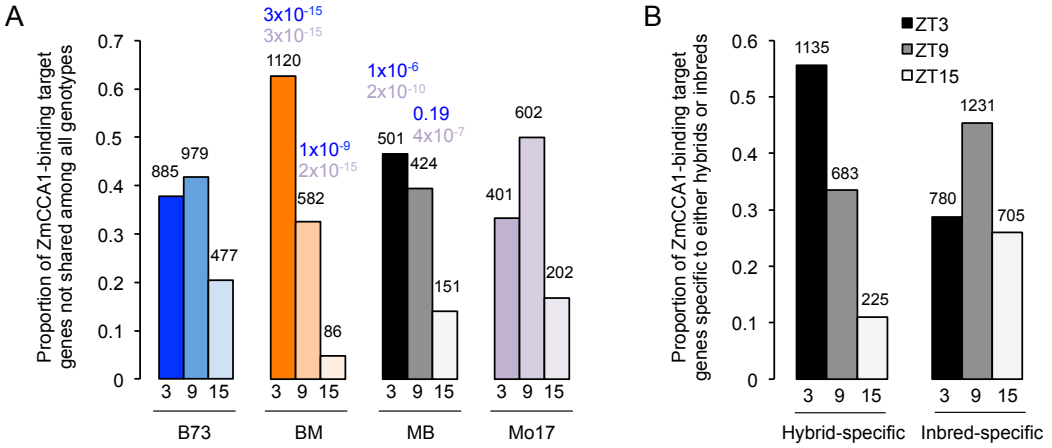

Figure 4.4 Temporal regulations of ZmCCA1-binding targets.

(A) Proportion of ZmCCA1-binding targets that were not shared. Statistical differences of ZT3 or ZT9 frequencies between the inbreds and F1 hybrids were calculated using Fisher's exact test. Blue and purple values indicate the comparison with B73 and Mo17, respectively. The number of peaks or genes is shown above each bar. (B) Proportion of ZmCCA1-binding targets that were specific to either hybrids or inbreds. Statistical differences of ZT3, ZT9 or ZT15 frequencies between the hybrid-specific and inbred-specific target genes were calculated using Fisher's exact test (ZT3, p-value = 2.2E-15; ZT9, p-value = 3.6E-15; ZT15, p-value = 2.9E-15). The number of genes is shown above each bar.

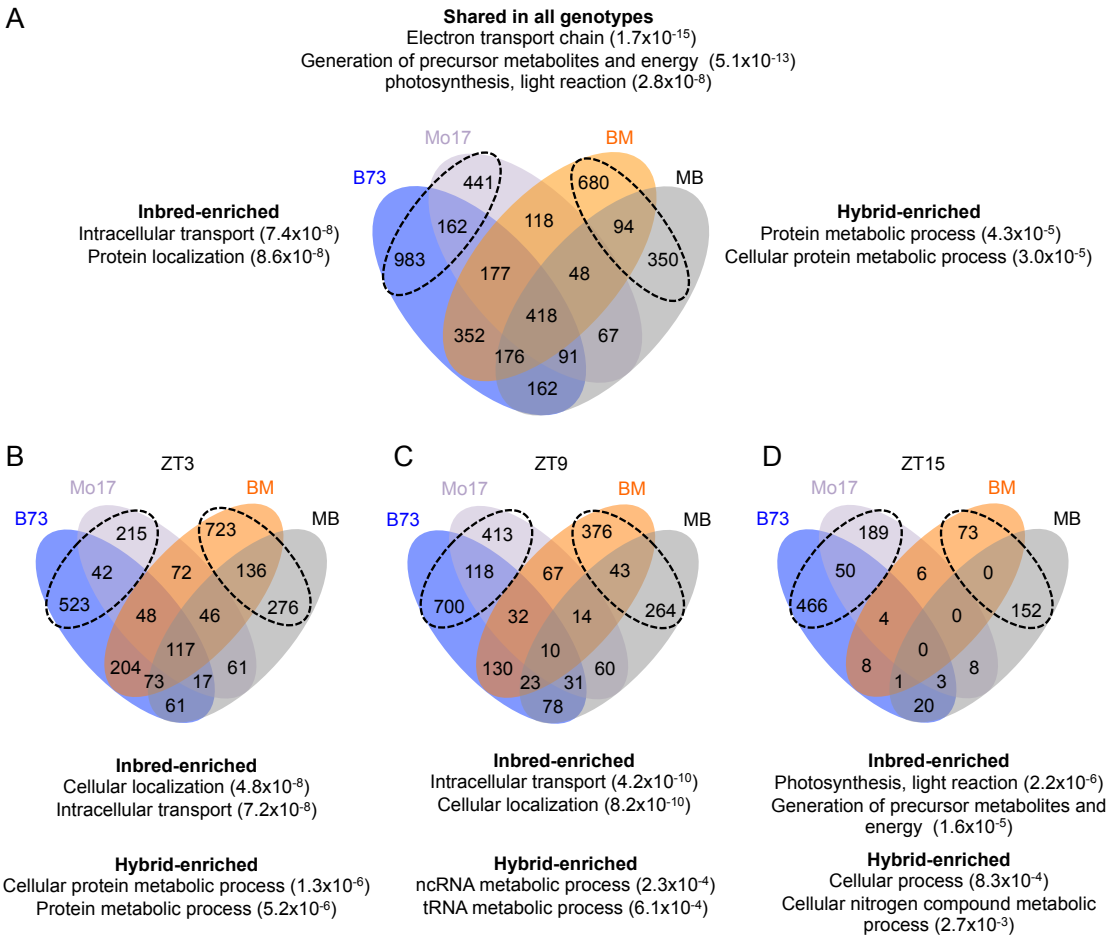

Figure 4.5 Diverse biological pathways are enriched in ZmCCA1s target genes.

(A) Venn diagram showing genotype-dependent and -shared ZmCCA1s targets in the inbreds and hybrids (B73, n = 2521; Mo17, n = 1522; BM, n = 2063; MB, n = 1406). (B-D) Venn diagram showing genotype-dependent and -shared ZmCCA1s targets at ZT3 (B), ZT9 (C) or ZT15 (D). Inbred-specific or hybrid-specific genes are indicated by dashed circles. Enriched functional pathways in GO analysis are shown with false-discovery rate adjusted p-value (Hypergeometric test).

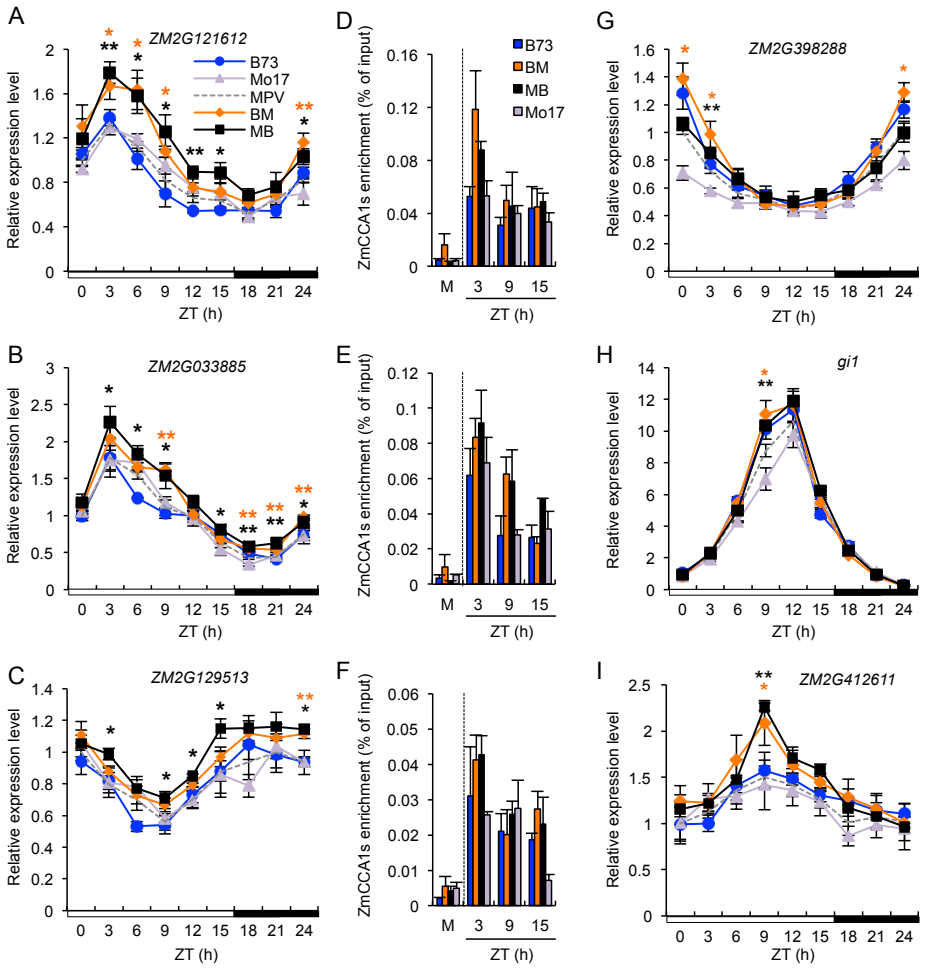

Figure 4.6 Diurnal expression levels of ZmCCA1b-bound carbon fixation genes in response to the phase-shift of ZmCCA1-binding in the hybrids.

(A-C and G) Relative expression levels (means ± SEM, n = 3) of the morning-phased carbon fixation genes, including *ZM2G121612* (A), *ZM2G033885* (B), *ZM2G129513* (C) and *ZM2G398288* (G), every 3 hours in a 24-hour period (light/dark cycle is shown below the histogram). The relative expression level in MPV at ZT0 was set to 1. Significant difference between hybrids and MPV was calculated using Student's t-test, *p-value < 0.05 and **p-value < 0.01 (D-F) ChIP-qPCR validation of ZmCCA1 ChIP enrichments on *ZM2G121612* (D), *ZM2G033885* (E) and *ZM2G129513* (F). M indicates the enrichment in mock samples at ZT3. The EMSA analysis confirmed ZmCCA1b-bining to the promoter of *ZM2G121612*, *ZM2G033885* and *ZM2G129513 in vitro* (S9 Fig). (H-I) Relative expression levels (means ± SEM, n = 3) of the afternoon-phased genes, *gi1* (H) and *ZM2G412611* (I), genes every 3 hours in a 24-hour period (light/dark cycle is shown below the histogram). The relative expression level in MPV at ZT0 was set to 1. Significant difference between hybrids and MPV was calculated using Student's t-test, *p-value < 0.05 and **p-value < 0.01.

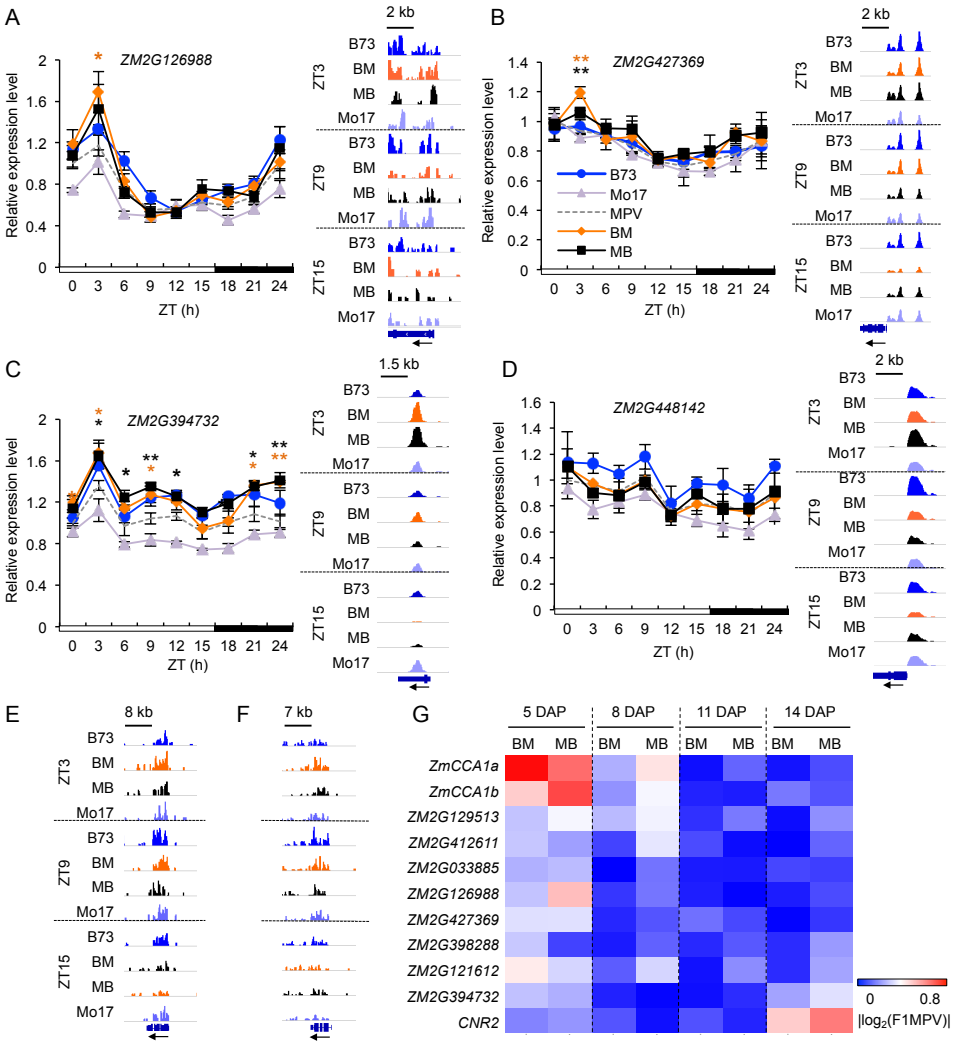

Figure 4.7 Temporal shift of ZmCCA1-binding target (carbon fixation) genes and their expression.

(A-D) Relative expression levels (means ± SEM, n = 3) of *ZM2G126988* (A), *ZM2G427369* (B), *ZM2G294732* (C), and *ZM2G448142* (D) every 3 hours in a 24-hour period (light/dark cycle is shown below the histogram). The relative expression level in MPV at ZT0 was set to 1. Significant difference between MPV and hybrids was calculated using Student's t-test, *p-value < 0.05 and **p-value < 0.01. The right panel for each gene shows ZmCCA1-binding peaks at ZT3, ZT9 and ZT15. The Y-axis indicates input-subtracted read density on a same-scale for all genotypes and time-points. Arrows indicate gene orientation. (E-F) ZmCCA1-binding enrichments of *ZM2G398288* (E) and *gi1* (F) at ZT3, ZT9 and ZT15. Notations are the same as above. (G) Heatmap of qRT-PCR data ($|\log_2(F1/MPV)|$, n = 3) showing nonadditive expression of maize *CCA1* and carbon fixation genes in the hybrids at ZT3 except for *ZM2G412611* at ZT9. *CNR2* was used as a marker for nonadditive expression in the seedlings.

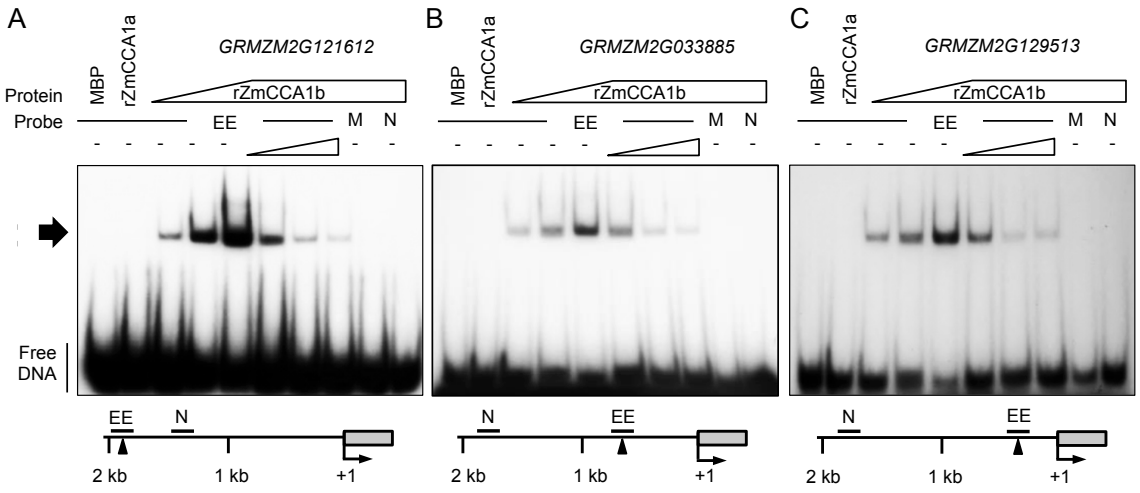

Figure 4.8 Verification of ZmCCA1-binding to promoters of carbon fixation genes *in vitro*.

Binding of recombinant ZmCCA1b (rZmCCA1b) to the promoters of ZmCCA1-binding target genes associated with carbon fixation *in vitro*. Radioisotope-labeled DNA probes (endogenous promoter fragments) were incubated in the presence of MBP (1 pmol), rZmCCA1a (1 pmol), and rZmCCA1b (0.5, 1 and 2 pmol). Shifted protein-DNA complexes are indicated by the arrow. Competitors: 25X, 50X and 100X molar excess of unlabeled prompter DNA. M: DNA in which EE or CBS site was mutated; N: no EE or CBS site in the DNA fragment. Location of probes for each gene is shown below the gel image. Arrowheads represent EE or CBS site. Numbers are relative to the transcription start site (+1) and 5" UTR (grey box).

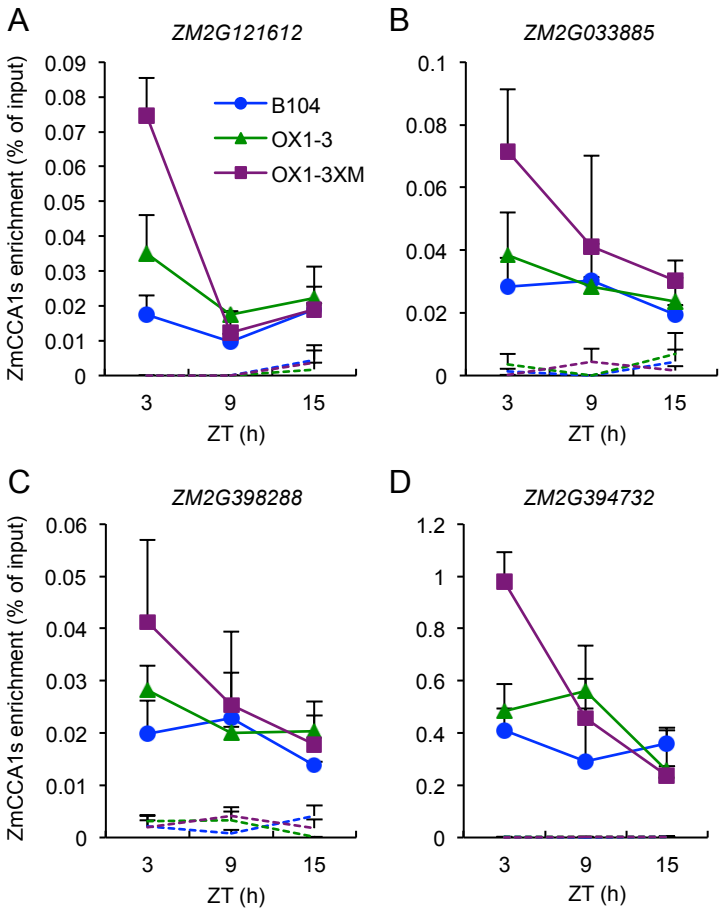

Figure 4.9 Temporal shift of ZmCCA1-binding to target (carbon fixation) genes in *ZmCCA1b* overexpression line (*OX1-3*) and F1 hybrids.

ChIP-qPCR assays were performed on ZmCCA1-binding target genes in B104, *OX1-3* and F1 hybrid (*OX1-3*XMo17) lines. ChIP-qPCR values are represented relative to the corresponding input values (means ± SEM, n = 2-3). Dash lines represent ChIP-qPCR data from control (mock) samples.

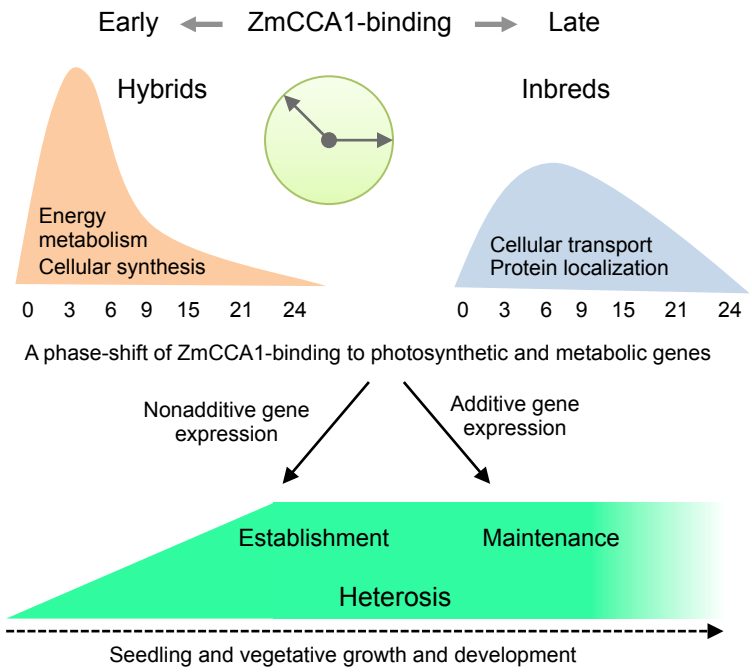

Figure 4.10 A phase-shift model for heterosis.

ZmCCA1-regulatory networks orchestrate multiple biological pathways for growth heterosis, which is established at early stage and subsequently maintained during the seedling development. The altered temporal binding of ZmCCA1s to carbon fixation genes in F1 hybrids, which causes nonadditive gene expression, increasing carbon fixation capacity and leading to heterosis. At the stage of maintenance, additive gene expression of the carbon fixation genes is predominant in the hybrids, suggesting a developmental coordination of additive and nonadditive gene expression for growth heterosis. Phase distributions of ZmCCA1-bindings are shown in orange for F1 hybrids and in blue for the inbreds.

Name	Gene ID	Forward sequence	Reverse sequence
ZmCCA1a	GRMZM2G474769	TTGGTGAGCCAAGGGCTTCCTTT	CTGTTCGGCGCAATTTCAGCTT
ZmCCA1b	*GRMZM2G014902*	CGAAGCATCCACATAATTGATTT	GCACTGCATTGCAAGATCTGA
gi1	*GRMZM2G107101*	GCAATACCAGCTCATTGGATAGTGT	CAACATCGCCATTCAGTAGGACCTC
ZmPRR59	*GRMZM2G135446*	TCTTGCTGCCTCCCAATGACCATA	ACTACTTGCACCAGCATCTTCCCT
ZmTOC1a	*GRMZM2G020081*	GCCAACCAATACGGATGTCAT	TGGATCGTCATCTTCGTCTTCA
-	*GRMZM2G121612*	GCTGGCGCAGACTTCATTCTTGTT	TGTGCTTGAGCCCGATCCTTATCA
-	*GRMZM2G033885*	ATCGACCGGAACGTGTTCTTGGAT	ATTACAAACCACCAAGTCCCAGCG
-	*GRMZM2G398288*	TCCAATGATGCAGTGGTTGGTA	AGCTGGCTTTGAAGGTGAATCT
-	*GRMZM2G427369*	TGGTCCACAGGGTTCATAACTA	AACAGTCGGACAAGTGGGTAAT
-	*GRMZM2G126988*	TAAGTGTCGCAAGGGTGGATGACA	TGGTACGAAGGGTTTCGTTGGTGA
-	*GRMZM2G129513*	CGTGTTCAGCATGCCATGCAGATC	TCTCAGCAAGCAATTCAGCTTCGC
-	*GRMZM2G394732*	GCATTAATTGCAACTTCCTCAG	CCAGAAAGACTAATCCAATCCA
-	*GRMZM2G412611*	TGTGGCAAAGAGTGTCAAGGGAGA	TTCCTTGCTCGGACTGACACATGA
-	*GRMZM2G448142*	TCTCGACAATACGAAGCAATAGG	TGCTTAGTGTGTGACTCGTTAG
CNR2	*GRMZM2G151230*	CGGCGGCGGCGGCTACTACCAG	GGCAGTCGTCGAAGCAGTTGCAGA
18srRNA	*AF168884*	TCTGTGATGCCCTTAGATGTTCTG	CTGTCGGCCAAGGCTATATACT

Table 4.1 List of primer used in qRT-PCR (5' to 3')

Name	Gene ID	Forward sequence	Reverse sequence
-	*GRMZM2G033885*	GGCAACCATGAAAACAATTGGTG	CCGGACTCACGAAGGTGAATGCA
-	*GRMZM2G129513*	GAGCACAGATCCCTTGGTAT	CATCCTAGCATTGATATCGT
-	*GRMZM2G121612*	GGTAGCCTGAAATCCTTATGCC	GAGCAACTAAACTCCAAAGATC
-	*GRMZM2G337113*	CAACCAAACATACAATCTATCTC	GCTGATATATTCACCAAACCTC
-	*GRMZM2G398288*	TCCAATGATGCAGTGGTTGGTA	AGCTGGCTTTGAAGGTGAATCT
-	*GRMZM2G394732*	ACCCAGGTGGATATACCCTATAA	GGAATCGAACCCGTATCTTCTC

Table 4.2 List of primer used in ChIP-qPCR (5' to 3')

Name	Gene ID	Description	Sequence
-	*GRMZM2G121612*	F	CTAAATTATAAGATATTTTATATTTTTTA
		R	TAAAAAATATAAAATATCTTATAATTTAG
		M_F	CTAAATTATAAtcgcggggATAcgagcgA
		M_R	TcgctcgTATccccgcgaTTATAATTTAG
		N_F	GTTCTTCTTCTTGCCACTCTTAGGTAGCG
		N_R	CGCTACCTAAGAGTGGCAAGAAGAAGAAC
-	*GRMZM2G033885*	F	TCAGCACTGAAAAAATATCTGAGGTTGAAA
		R	TTTCAACCTCAGATATTTTTCAGTGCTGA
		M_F	TCAGCACTGAccccgcgaTGAGGTTGAAA
		M_R	TTTCAACCTCAtcgcggggTCAGTGCTGA
		N_F	AGGTAGTGCTTGGTATTGACATCAGACTT
		N_R	AAGTCTGATGTCAATACCAAGCACTACCT
-	*GRMZM2G129513*	F	ATTAAATATAAAAATATCTAAATTCATAA
		R	TTATGAATTTAGATATTTTTATATTTAAT
		M_F	ATTAAATATAccccgcgaTAAATTCATAA
		M_R	TTATGAATTTAtcgcggggTATATTTAAT
		N_F	CGCGGTGAAGCATCACTGGGCCCAAGGGG
		N_R	CCCCTTGGGCCCAGTGATGCTTCACCGCG

Table 4.3 List of oligonucleotides used in EMSA (5' to 3')

EE or CBS elements are underlined. Mutated sequences are lower cases. F: forward strand; R: reverse strand; M: DNA in which EE or CBS site was mutated; N: no EE or CBS site in the DNA fragment.

Chapter 5. An improved protocol for chromatin immunoprecipitation (ChIP) in maize and library construction for deep sequencing (ChIP-seq)

5.1 Introduction

Many complex developmental, growth and differentiation processes in eukaryotic systems are mediated by dynamic protein-DNA interactions that direct transcriptional regulations. One of best examples is transcriptional regulation by transcription factors (TFs), which is considered as the primary level of gene regulations [119]. Interactions of DNA with chromatin-associated proteins, such as chromatin remodelers and histone modifiers, are vital in epigenetic regulation of gene expression [31, 167, 168]. The main tool for mapping sites of these interactions *in vivo* is chromatin immunoprecipitation (ChIP), which were initially developed in cultured *Drosophila* cells [169] and has widely applied to other eukaryote species [170-173]. In short, proteins bound to the DNA are crosslinked *in vivo* by fixates such as formaldehyde. Specific protein-DNA complexes are then immunoprecipitated and purified by using a specific antibody against the DNA-binding protein of interest or a tag that is fused with the protein. In recent years, with increasing access to and decreasing costs of next-generation sequencing technology, ChIP followed by deep sequencing (ChIP-seq) has emerged as a powerful and popular tool for studying these protein-DNA interactions on a genome-wide scale in real time [120, 122, 174, 175]. Not surprisingly, since the recent completion of reference genome sequences for many crop species [103, 176-179], ChIP-seq technique has received a high demand from the research community [180-182]. However, it is still challenging and

time-consuming to prepare good-quality ChIP-seq library for their tissue materials, which is crucial to achieve a high signal-to-noise ratio in ChIP-seq data analysis. It is partly due to the technical difficulties in ChIP procedures for these crop plants, which have more complex cellular properties than *Arabidopsis* [55, 183, 184]. The relatively high cost for commercial kits of ChIP-seq library construction also remains a serious limitation to broader application of ChIP-seq technique to large-scale experiments.

Here, I present, to my knowledge, the first detailed explanation of a robust, time-efficient and low-cost protocol for a combination of ChIP and ChIP-seq library construction, which optimizes and integrates the existing ChIP and ChIP-seq library construction protocols [185-187] for maize. I have recently used this protocol for studying global DNA-binding profiles of CCA1 homologs in maize hybrids and inbred lines, and successfully generated reliable results (Figure 4.4; Ko *et al*., in revision). I have also used this protocol for investigating interactions of DNA with chromatin-associated proteins.

Basic overview of ChIP and ChIP-seq library construction protocol

Although many improvements and optimizations have been introduced [170-173, 186, 187] since its development in *Drosophila* more than two decades ago [169], the basic principle of ChIP remains the same. It consists of three basic steps, each of which has key procedures, as described in Figure 5.1A; tissue crosslinking, chromatin immunoprecipitation and DNA purification. Importantly, these steps are highly desirable to be modified for plant tissue materials because of their structural and anatomical uniqueness [173, 186], which makes ChIP more difficult in plants. Briefly, after the

tissue collection, the fresh plant material is immediately fixed to crosslink protein-DNA interactions using a fixative, formaldehyde in this protocol. Since the fixative needs to penetrate the plant cells passing through the waxy and thick cell wall, substantial amount and duration of vacuum are required until the plant tissue appears translucent. After the fixation, chromatin DNA is extracted, shared and then immunoprecipitated with an antibody against the DNA-bound protein of interest. The DNA is isolated by reverse crosslinking and purified. The purified DNA can directly be used for ChIP-seq library construction.

The workflow of ChIP-seq library construction for Illumina sequencing platforms consists of end repair, A-tailing, adaptor ligation and PCR enrichment of adaptor-ligated DNA according to standard New England BioLabs (NEB) and Illumina library construction procedures (Figure 5.1A). The purified ChIP DNA is subjected to end repair and A-tailing, which allow for successful ligation with adaptors. Following size selection using various methods including AMPure XP beads in this protocol, the DNA is enriched by PCR amplification with a low cycle (e.g. less than 20). The resulting library is immediately ready for deep sequencing.

Development of the protocol

I initially tested published *Arabidopsis* protocols for ChIP on maize tissue, but subsequently failed to have a satisfying enrichment of target DNA. I assumed that the complex cellular composition and secondary compounds of the maize tissue primarily impede penetrance of the fixative into maize tissue [55, 183, 184]. Since different types of tissue material require substantial optimizations of the procedures systematically, I

gathered pieces of the updated information from published protocols [185-187] and introduced crucial modifications into them, which finally allowed me to have reliable results.

Although ChIP-seq library construction kits for Illumina sequencing platforms are commercially available, the workflow is quite simple and appears not to require the relatively expensive kits. Importantly, the cost of using the kits incredibly goes up with a large amount of samples in the experimental set-up (Figure 5.1B). For this reasons, I started reconstructing the procedures with customized reagents based on the existing workflow. After confirming successful results from each ChIP and ChIP-seq library construction, I combined those into a single pipeline to generate a robust, time-efficient and low-cost protocol for my own studies.

Comparison with other methods

A distinctive feature of our protocol compared to others is the crosslinking condition. Since maize tissue has complex cellular properties [55, 183, 184], crosslinking buffer contents and vacuum time should be optimized to ensure penetration of the fixative, formaldehyde in this protocol. For example, formaldehyde concentration is inevitably higher and vacuum time is longer than in *Arabidopsis* protocols [185, 186]. Despite the simplicity of the components, the optimization of crosslinking condition was quite challenging because too much concentration of formaldehyde or longer time of vacuum may cause strong crosslinking and thus impede reverse crosslinking in later steps. Through a number of test runs, I noticed that vacuum system (e.g. different vacuum chambers) could also affect the efficiency of crosslinking. I used a side arm flask (125

ml) that is directly connected to a built-in vacuum system, which is likely available in most laboratories, via a rubber braided air hose (~60 cm) for each sample to maximize the efficiency of crosslinking. This standardized information for the crosslinking condition will help researchers generate consistent results.

Chromatin extraction is one of difficult steps especially in maize tissue owing to the high contents of secondary compounds such as polysaccharides [183, 184]. Therefore, I tested and integrated a number of improvements in buffer compositions and conditions made from published protocols [185-188], which allows for rapid extraction of good-quality crosslinked chromatin. Chromatin fragmentation is largely achieved by sonication or nuclease, each of which has own advantage and disadvantage [189]. This protocol uses sonication in Bioruptor system, which is one of most popular sonicators, with optimized reaction cycles effectively to yield DNAs in 0.2-2-kb range. Following the DNA purification, ChIP DNA is directly subjected to ChIP-seq library construction (Figure 5.1A). The basic workflow of our ChIP-seq library construction protocol is based on the commercial kits, yet it takes shorter and cheaper than buying those (Figure 5.1B). Therefore, this protocol for ChIP and ChIP-seq library in maize tissue is robust, time-efficient and cost-effective.

Advantages and disadvantages of this protocol

The first advantage of this protocol is its application to other grasses, of which cellular properties are similar with maize, such as sorghum and barley. Given the high demand of ChIP-seq in these species now and in the near future [181, 190], this protocol will serve as a good starting point to generate high-quality ChIP-seq library not only in

maize but also in those. Second, this protocol has already demonstrated to show reliable results using a native antibody against TFs (Figure 4.4; Ko *et al*., in revision). A widely used approach for strong ChIP enrichments in model systems is to generate stable transgenic lines, in which a tag is fused with proteins of interest, and then to immunoprecipiate TF-bound DNAs through the tag. However, generation of transgenic lines is, more often than not, problematic or time-consuming in crop species such as maize. In this regard, this protocol allows strong enrichments of protein-DNA interaction to be detected in maize using a native antibody as shown in my study (see Chapter 4). Lastly, although this protocol for ChIP-seq library construction does not diverge significantly from the standard commercial kits, it offers a significant benefit on the cost especially for large-scale ChIP-seq experiments (Figure 5.1B).

A disadvantage of this protocol is that it has not been validated in other tissue types (e.g. root and ear) than aerial tissue, which might have different cellular properties. When those tissues are used, modifications particularly in crosslinking and chromatin extraction will be needed to optimize the condition. As a limitation of all ChIP experiment in my protocol and others, the quality of the resulting library heavily relies on the specificity of the antibody used. A highly specific antibody will offer a high-level of enrichment relative to the background. Therefore, antibody characterization (e.g. immunoblot analyses) is strongly recommended for successful outcomes before preceding our protocol [191].

Experimental design

Tissue collection. Maize seedling tissue (whole aerial tissue) needs to be diced in ~1 cm using scissors to ensure successful penetration of formaldehyde. I recommend pooling at least two plants for a biological replicate to minimize biological variations. While *Arabidopsis* plants are generally grown in a growth chamber or growth room in the laboratory, crop plants such as maize seedlings are often grown in a greenhouse separated from the laboratory. If this is the case, crosslinking buffer without formaldehyde (Crosslinking buffer A) is initially used for the tissue collection in greenhouse. Then, Crosslinking buffer A is replaced by one with formaldehyde (Crosslinking buffer B) in the laboratory, right before vacuum is applied. I had ~10 min time interval of the collected tissue in Crosslinking buffer A with no problem.

Crosslinking. While sufficient crosslinking is indispensible in ChIP, excessive crosslinking adversely affects DNA enrichment; it may reduce antibody accessibility to antigen, efficiency of chromatin fragmentation, DNA isolation or reverse crosslinking [192]. Therefore, the extent of crosslinking should empirically be optimized, which includes tissue size, time, formaldehyde concentration and temperature. Through a number of test runs and from published studies [187, 188], I have learned that 3% formaldehyde at room temperature during 20 min is optimal for a successful crosslinking in maize tissue in my experimental setting.

Antibody. The antibody is the most crucial factor for a successful ChIP and ChIP-seq analyses [186, 191]. The specificity of antibody, therefore, should be verified using suggested methods [191] before a ChIP. Additionally, motif analysis can be performed in binding peaks to validate for specificity of the antibody in ChIP-seq data analysis, if the

protein of interest is a well-characterized TF. The binding specificity to the motif in ChIP-seq can be further validated using independent methods (e.g. EMSA *in vitro*). The antibody concentration for ChIP must be empirically optimized through test runs to obtain a high-level enrichment of DNA. For TF ChIP or ChIP-seq, the antibody can be developed to recognize the endogenous target protein or a so-called tag fused with the target protein. Either way has own advantage and disadvantage. For example, use of a native antibody greatly reduces the risk of artifacts that may be generated using a tag antibody (e.g. unexpected functions caused by the tagged protein or the absence of the endogenous protein). In the other hand, a tag antibody results in higher specificity, namely high signal-to-noise ratio, which may not be achieved using a native antibody. Tagging also provides the experimental options for tissue- or cell-specific protein-DNA interactions while it often takes years to generate a stable transgenic line in crops such as maize.

Design of controls. To ensure the success of ChIP, both input (sonicated chromatin) and mock (IP with no antibody under the same condition) samples need to be included in the entire process. The input will be indicative for the existence of the DNA, serving a positive control for the entire ChIP. The mock may provide a background level for ChIP, since it is carried out the same way as ChIP samples. After ChIP, yet before ChIP-seq library construction, qPCR analysis is routinely performed in ChIP samples, including both controls, using positive (e.g. known binding sites) and negative (e.g. unrelated loci) gene primer sets in order to test the enrichment of DNA. In the qPCR analysis, signal from the input sample can be used for data normalization (known as percent of input

method) [193] while one from the mock sample indicates the background level. Importantly, for ChIP-seq analysis, the input sample needs to be sequenced along with the ChIP sample and used as a reference for searching ChIP-enriched regions (peaks) in peak-calling algorithms. Although the mock sample can also be used for the reference in theory, it normally has insufficient amount of DNA for the library construction and, thus, may be noisy or biased.

5.2 Materials

REAGENTS

Plant material: fresh maize tissue (e.g. whole aerial tissue of seedling), 1-3 g

Sucrose (Sigma-Aldrich, cat. no. S5016)

Formaldehyde 37% (vol/vol) (Sigma-Aldrich, cat. no. F8775) CAUTION Toxic, handle it with care. Avoid contact with body.

NaCl (Sigma-Aldrich, cat. no. 31434)

EDTA disodium salt (Fisher Scientific, cat. no. BP120-500)

Glycine (Fisher Scientific, cat. no. 12007-0050)

Tris base (Fisher Scientific, cat. no. BP152-5)

Sodium dodecyl sulfate (SDS; Sigma-Aldrich, cat. no. S5881) CAUTION Carcinogenic, handle it with care. Avoid contact with body.

Potassium chloride (KCl; EM, cat. no. PX1405.5)

Magnesium chloride ($MgCl_2$; Fisher Scientific, cat. no. M33-500)

Lithium chloride (LiCl; Fisher Scientific, cat. no. L121-500)

Sodium acetate (NaOAc; Fisher Scientific, cat. no. S210-500)

Sodium bicarbonate ($NaHCO_3$; Sigma-Aldrich, cat. no. S6297)

Sodium deoxycholate (Deoxycholic acid, sodium salt) (Fisher Scientific, cat. no. BP349-100)

Calcium chloride ($CaCl_2$; Fisher Scientific, cat. no. C614-500)

Glycogen, 20 μg/μL (Invitrogen, cat. no. 10814-010)

Phenol/chloroform/isoamyl alcohol (25:24:1) (Fisher Scientific, cat. no. BP1752-400) CAUTION Toxic, handle it with care. Avoid contact with body.

PMSF (Sigma-Aldrich, cat. no. P7626)

HEPES (Fisher Scientific, cat. no. BP310-100)

Triton X-100 (Sigma-Aldrich, cat. no. T8787)

NP-30 detergent solution (USBiological, cat. no. N3500)

Ultrapure distilled water (Invitrogen, cat. no. 10977-015)

β-Mercaptoethanol, 14.2 M (Sigma-Aldrich, cat. no. M3148) CAUTION Toxic, handle it with care. Avoid contact with body.

Salmon sperm DNA/ Protein A agarose beads (EMD Millopore, cat. no. 16-157)

1 mg/ml Pepstatin A (Sigma-Aldrich, cat. no. P5318)

Protease inhibitor cocktail (Sigma-Aldrich, cat. no. P2714)

Proteinase K (Fisher Scientific, cat. no. BP1700)

Agarose (Sigma-Aldrich, cat. no. A0169)

Ethidium bromide (Fisher Scientific, cat. no. BP1302-10) CAUTION Powerful mutagen, handle it with care. Avoid contact with body.

Ethanol, 95% (vol/vol) (PHARMCO-AAPER, cat. no. 111000200)

Isopropanol (Fisher Scientific, cat. no. A416-4)

Deoxyadenosine 5′- Triphosphate (dATP; New England BioLabs, cat. no. N0440S)

Deoxynucleotide Solution Mix, 10 mM each dNTP (Promega, cat. no. U1515)

NEB buffer 2 (New England BioLabs, cat. no. B7002S)

T4 DNA Polymerase (New England BioLabs, cat. no. M0203S)

DNA Polymerase I, Large Fragment (Klenow; New England BioLabs, cat. no. M0212S)

T4 Polynucleotide Kinase (PNK; New England BioLabs, cat. no. M0201S)

Quick T4 DNA Ligase (New England BioLabs, cat. no. M2200S)

NEBNext Multiplex Oligos for Illumina (New England BioLabs, cat. no. E7335)

NEBNext High-Fidelity 2X PCR Master Mix (New England BioLabs, cat. no. M0541)

AMPure XP beads (Beckman Coulter, cat. no. A63881)

FastStart Universal SYBR Green Master (ROX) (Roche Applied Science, cat. no. 04913914001)

EQUIPMENT

Aluminum foil roll (BROADWALK, cat. no. BWK7124)

Benchtop centrifuge (Beckman Coulter, cat. no. Allegra X-12R)

Bioanalyer (Agilent, cat. no. 2100)

Bioruptor TPX microtubes, 1.5 ml (Diagenode, cat. no. C20010010-50)

Bioruptor sonicator (Diagenode, cat. no. UCD-200)

Cold room, -4 °C

Conical tubes, 50 ml (greiner bio-one, cat. no. 227 261)

Conical tubes, 15 ml (greiner bio-one, cat. no. 188 271)

Electrophoresis system (Fisher Scientific, cat. no. FB-SB-710)

Freezers, -80 °C and -20 °C

Ice bucket (Sigma-Aldrich, cat. no. Z220531)

Liquid nitrogen (LN_2)

Magnetic particle concentrator (Invitrogen, cat. no. 123.21D)

Microcentrifuge tubes, 1.7 ml (AXYGEN, cat. no. 311-05-051)

Microcentrifuge tubes, 2 ml (BioExpress, cat. no. C-3219-1)

Low-adhesion microcentrifuge tubes, 1.7 ml (AXYGEN, cat. no. 311-05-051)

Miracloth (CALBIOCHEM, cat. no. 475855)

NanoDrop (NanoDrop, cat. no. ND1000)

PCR machine (Applied Biosystems, cat. no. 2720)

Real-time PCR machine (Roche Applied Science, LightCycler 96 System)

LightCycler 480 multiwell plate 96 (Roche Applied Science, cat. no. 05 102 413 001)

LightCycler 480 sealing foil (Roche Applied Science, cat. no. 04 729 757 001)

Standard PCR strips, 200 µl (BioExpress, cat. no. T-3035)

Paper towels (Georgia-Pacific, cat. no. 10073310233045)

Portable LN_2 container (NALGENE, cat. no. 4150-1000)

Refrigerated microcentrifuge (Eppendorf, cat. no. 5415R)

Rotator (LABQUAKE, cat. no. 3.625.485)

Side arm flasks, 125 ml (PYREX, cat. no. 5360)

Green neoprene stoppers (Refinery Supply, cat. no. 18155-001)

Rubber braided air hose, 60 cm

Thermo mixer (Eppendorf, cat. no. 5355 22246)

Vacuum system (Built-in)

Vortexer (Scientific Industries, cat. no. G560)

REAGENT SETUP

Pepstatin A, 1 mg/ml. Dissolve at 1 mg/ml in 10% (v/v) acetic acid in methanol (9:1 methanol:acetic acid). Stock solutions at 1 mg/ml are stable at least a week at 4 °C. If solutions become darker yellow, the reagent is hydrolyzing.

Protease inhibitor cocktail, 10X. Add 10 ml ddH_2O to a new bottle (Sigma, P2714-1BTL) and mix. Make aliquots and store at -20 °C.

PMSF, 100 mM. Dissolve 1.74 g in 100 ml isopropanol and store at 4 °C.

Crosslinking buffer A. Add 80 ml Sucrose (2 M stock), 4 ml Tris-HCl pH 8.0 (1 M stock), 4 ml PMSF (0.1 M stock) and 349.8 ml ddH_2O by a final volume of 400 ml. CRITICAL STEP Make fresh buffer from stock solutions just before use. It may be possible to make a premix without the protease inhibitor (PMSF) in advance and add PMSF to the premix directly before use in case that many samples need to be harvested in a time series experiment. It is recommended to distribute the buffer to the side arm flaks before harvesting tissues to save any interval time delay. Final concentration: 0.4 M Sucrose, 10 mM Tris-HCl pH 8.0 and 1 mM PMSF.

Crosslinking buffer B. Add 100 ml Sucrose (2 M stock), 5 ml Tris-HCl pH 8.0 (1 M stock) and 5 ml PMSF (0.1 M stock), 40 ml formaldehyde (37% stock), 176 µl β-Mercaptoethanol (14.2 M stock) and 312 ml ddH_2O by a final volume of 500 ml.

CRITICAL STEP Make fresh buffer from stock solutions just before harvesting tissues and keep it on ice until use. It may be possible to make premix without the protease inhibitor (PMSF) in advance and add PMSF to the premix directly before using the buffer in case that many samples need to be harvested in a time series experiment. Final concentration: 400 mM Sucrose, 10 mM Tris-HCl pH 8.0, 1 mM PMSF, 3% Formaldehyde and 5 mM β-Mercaptoethanol.

Nuclei isolation buffer A. Add 26 ml Sucrose (2 M stock), 1.3 ml Tris-HCl pH 8.0 (1 M stock) and 1.3 ml PMSF (0.1 M stock), 45.8 µl β-Mercaptoethanol (14.2 M stock), 25 µl Protease inhibitor cocktail (10X) and 201.4 ml ddH_2O by a final volume of 500 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it on ice until use. Final concentration: 400 mM Sucrose, 10 mM Tris-HCl pH 8, 1 mM PMSF, 5 mM β-Mercaptoethanol and the manufacture's instruction for Protease inhibitor cocktail (25 µl 10X Protease inhibitor per 5 ml final volume)

Nuclei isolation buffer B. Add 500 µl Sucrose (2 M stock), 100 µl MgCl2 (1 M stock), 500 µl Triton X-100 (20% stock), 100 µl Tris-HCl pH 8.0 (1 M stock), 100 µl PMSF (0.1 M stock), 3.4 µl β-Mercaptoethanol (14.2 M stock), 50 µl Protease inhibitor cocktail (10X) and 8.6466 ml ddH_2O by a final volume of 10 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it on ice until use. Final concentration: 0.25 M Sucrose, 10 mM MgCl2, 1% Triton X-100, 10 mM Tris-HCl pH 8, 1 mM PMSF, 5 mM β-Mercaptoethanol and the manufacture's instruction for Protease inhibitor cocktail (25 µl 10X Protease inhibitor per 5 ml final volume).

Nuclei isolation buffer C. Add 17 ml Sucrose (2 M stock), 40 ml $MgCl_2$ (1 M stock), 150 µl Triton X-100 (20% stock), 200 µl Tris-HCl pH 8.0 (1 M stock), 200 µl PMSF (0.1 M stock), 7 µl β-Mercaptoethanol (14.2 M stock), 100 µl Protease inhibitor cocktail (10X) and 2.303 ml ddH_2O by a final volume of 20 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it on ice until use. Final concentration: 1.7 M Sucrose, 2 mM MgCl2, 0.15% Triton X-100, 10 mM Tris-HCl pH 8, 1 mM PMSF, 5 mM β-Mercaptoethanol and the manufacture's instruction for Protease inhibitor cocktail (25 µl 10X Protease inhibitor per 5 ml final volume).

Nuclei lysis buffer. Add 1 ml HEPES pH 7.5 (500 mM stock), 300 µl NaCl (5 M stock), 20 µl EDTA (0.5 M stock), 100 µl PMSF (100 mM stock), 500 µl SDS (20% stock), 500 µl Na deoxycholate (10% stock), 500 µl Triton X-100 (20% stock) and 10 µl Pepstatin A (1 mg/ml stock), 50 µl Protease inhibitor cocktail (10X) and 7.42 ml ddH_2O by a final volume of 10 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it on ice until use. Final concentration: 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM PMSF, 1% SDS, 0.1% Na deoxycholate, 1% Triton X-100 and the manufacture's instruction for Protease inhibitor cocktail (25 µl 10X Protease inhibitor per 5 ml final volume).

ChIP dilution buffer for washing beads. Add 5 ml HEPES pH 7.5 (500 mM stock), 1.5 ml NaCl (5 M stock), 100 µl EDTA (500 mM stock), 2.5 ml Triton X-100 (20% stock) and 40.9 ml ddH_2O by a final volume of 50 ml. Although it is recommend to make buffer freshly, it is may be possible to make it a week in advance and store at -4 °C. Final

concentration: 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA and 1% Triton X-100.

ChIP dilution buffer. Add 5 ml HEPES pH 7.5 (500 mM stock), 360 µl NaCl (5 M stock), 24 µl EDTA (500 mM stock), 120 µl PMSF (100 mM stock), 120 µl Na deoxycholate (10% stock), 600 µl Triton X-100 (20% stock), 12 µl Pepstatin A (1 mg/ml stock), 1.2 ml Protease inhibitor cocktail (10X) and 8.364 ml ddH_2O by a final volume of 12 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it on ice until use. Final concentration: 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM PMSF, 0.1% Na deoxycholate, 1% Triton X-100, 1 µg/ml Pepstatin A and the manufacture's instruction for Protease inhibitor cocktail (25 µl 10X Protease inhibitor per 5 ml final volume).

Pre-equilibrated salmon sperm DNA/protein A agarose beads. Take two each 550 µl well-mixed salmon sperm DNA/protein A agarose beads in a 1.7 ml microcentrifuge tube separately, centrifuge these two tubes at 3,800*g* for 2 min at 4 °C to pellet the beads, discard the supernatant of each tube, add 1 ml of dilution buffer for washing beads, mix for 2 min at 4 °C with gentle rotation, centrifuge at 3,800*g* for 2 min at 4 °C to pellet the beads, and repeat the washing. After the 2nd wash, resuspend beads in 650 µl ChIP dilution buffer for each tube and equally distribute pre-equilibrated slurry of each tube into 8 microcentrifuge tubes (~110 µl/tube). Now, there are 16 microcentrifuge tubes containing ~110 µl pre-equilibrated slurry.

Low salt wash buffer. Add 750 µl NaCl (5 M stock), 250 µl SDS (20% stock), 625 µl Triton X-100 (20% stock), 100 µl EDTA (500 mM stock) and 500 µl Tris-HCl pH 8.0 (1

M stock) and 22.775 ml ddH_2O by a final volume of 25 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it on ice until the use. Final concentration: 150 mM NaCl, 0.2% SDS, 0.5% Triton X-100, 2 mM EDTA and 20 mM Tris-HCl pH 8.0.

High salt wash buffer. Add 2.5 ml NaCl (5 M stock), 250 µl SDS (20% stock), 625 µl Triton X-100 (20% stock), 100 µl EDTA (500 mM stock) and 500 µl Tris-HCl pH 8.0 (1 M stock) and 21.025 ml ddH_2O by a final volume of 25 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it on ice until use. Final concentration: 500 mM NaCl, 0.2% SDS, 0.5% Triton X-100, 2 mM EDTA and 20 mM Tris-HCl pH 8.0.

LiCl wash buffer. Add 6.25 ml LiCl (1 M stock), 2.5 ml Na deoxycholate (10% stock), 250 µl Tris-HCl pH 8.0 (1 M stock), 250 µl NP-40 (100% stock), 50 µl EDTA (500 mM stock) and 15.7 ml ddH_2O by a final volume of 25 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it on ice until use. Final concentration: 250 mM LiCl, 1% Na deoxycholate, 10 mM Tris-HCl pH 8.0, 1% NP-40 and 1 mM EDTA.

TE buffer. Add 250 µl Tris-HCl pH 8.0 (1 M stock), 50 µl EDTA (500 mM stock) and 24.7 ml ddH_2O by a final volume of 25 ml. The buffer is stable for months at room temperature. Pre-cool it on ice until use.

Elution buffer. Add 300 µl SDS (20% stock), 100 mg Sodium bicarbonate and 11.7 ml ddH_2O by a final volume of 12 ml. CRITICAL STEP Prepare fresh buffer from stock solutions and keep it at room temperature to prevent SDS precipitation.

5.3 Procedures (designed to handle 8 samples each for ChIP or for ChIP-seq library construction)

Tissue collection TIMING 10-15 min

1. Harvest ~3 g maize tissue (whole aerial or leaves) and cut by ~1 cm size using a scissor. Collect cut tissue fragments into a 125 ml side arm flask containing 30 ml crosslinking buffer A. CRITICAL STEP It is highly recommended to have at least two biological replicates from pools of multiple plants for ChIP-seq analysis.

Crosslinking TIMING 40-45 min

2. Drain crosslinking buffer A, and add 40 ml crosslinking buffer B to the flask.

3. Apply vacuum to the flask, containing cut tissue fragments and crosslinking buffer B, at room temperature during 20 min. CRITICAL STEP Vacuum time could be slightly different depending on the vacuum system one uses, so it is recommended to have test runs before the experiment. It is also important to submerge the tissues completely for a successful vacuum infiltration by gently shaking the flask. TROUBLESHOOTING

4. Stop vacuum, add 2.5 ml 2 M Glycine (final concentration 125 mM), mix well by gently shaking the flask several times and apply vacuum during 5 min at room temperature.

5. Drain crosslinking buffer B, and wash the tissues quickly with pre-cooled ddH_2O.

6. Dry the tissues by gently blotting between paper towels, then put them in an aluminum foil bag with appropriate labels and quick-freeze in liquid nitrogen. PAUSE POINT The tissue can be stored at -80 °C for months to years.

Nuclei isolation and chromatin fragmentation TIMING 6 h

7. Grind the tissues to a fine powder in liquid nitrogen using pre-chilled motor and pestle. CRITICAL STEP The tissue needs to be ground as a fine powder for successful nuclei isolation. The tissues should not be thawed during the grinding.

8. Transfer the powder to a pre-chilled 50 ml conical tube, and add 30 ml pre-chilled Nuclei isolation buffer A to the tube. Vortex the tube shortly, and rotate it for 30 min at 4 °C. CRITICAL STEP It is important to achieve a complete homogenization.

9. Filter the completely homogenized slurry through Miracloth using a funnel into a new 50 ml conical tube, and centrifuge the filtrate at 2,880*g* for 10 min at 4 °C.

10. Discard the supernatant, resuspend the pellet in 1 ml pre-chilled Nuclei isolation buffer B by pipetting, transfer the solution to a 1.7 ml microcentrifuge tube and centrifuge the filtrate at 12,000*g* for 10 min at 4 °C.

11. Discard the supernatant and resuspend the pellet in pre-chilled 300 µl Nuclei isolation buffer C using a sterilized toothpick.

12. Add 1.5 ml Nuclei isolation buffer C to a new 2 ml microcentrifuge tube, overlay the 300 µl nuclei lysate top of it, and centrifuge at 16,000*g* for 60 min at 4 °C.

13. Discard the supernatant and resuspend the pellet in 310 µl pre-chilled Nuclei lysis buffer using a sterilized toothpick. Take 10 µl sample and store on ice. This will serve as a control for sonication efficiency test.

14. Sonicate the nuclei suspension in 10 cycles of 30 sec-on and 30 sec-off with high power using Bioruptor sonicator. CRITICAL STEP The sonication must be performed in cold condition to avoid overheating the samples.

15. Centrifuge samples at 13,800*g* for 10 min, 4 °C.

16. Transfer the supernatant to a new 1.7 ml microcentrifuge tube. Take 10 µl of the supernatant to check sonication efficiency test (Step 17). The remaining supernatant is ready to be used for pre-clearing or can be stored at -80 °C until use. PAUSE POINT At this point, sonicated chromatin can be stored at -80 °C for ~3 months.

17. Mix chromatin samples (from Step 14 and 17) with 10 µl ddH_2O and 4 µl 40% (wt/vol) Sucrose. Run 12 µl of each mixture on a 1.5% (wt/vol) agarose gel in 1X TAE buffer at 105 V for 40 min. The sample from Step 16 should have a smear from 0.2 to 2 kb, but concentrated ~500 bp compared to one from Step 13. TROUBLESHOOTING

Immunoprecipitation TIMING 18-19 h

18. Save 50 µl of the supernatant (Step 17) as an input control and store it at -80 °C until use. Dilute the remaining 200 µl with 1,000 µl dilution buffer in a new 1.7 ml microcentrifuge tube. This is for 2 IPs; TF and mock.

19. Pre-clear the diluted chromatin by adding 70 µl pre-equilibrated salmon sperm DNA/protein A agarose beads (see REAGENT SETUP) and incubate for 1 h at 4 °C on a rotating wheel.

20. Centrifuge at 3,800*g* for 2 min at 4 °C, and equally divide pre-cleared diluted chromatin supernatant in a new 1.7 ml microcentrifuge tube, resulting 2 tubes of 600 µl.

21. Add 1-10 µg of an appropriate antibody to one of pre-cleared diluted samples, and the other sample will be treated in the sample way but without an antibody as a mock control. Incubate both samples overnight (18 h) at 4 °C on a rotating wheel. CRITICAL STEP The amount of antibody needs to be optimized through test runs.

Washing and elution TIMING 4 h

22. Add 70 µl pre-equilibrated salmon sperm DNA/protein A agarose beads to each sample and incubate it for 2 h at 4 °C on a rotating wheel.

23. Centrifuge at 3,800*g* for 2 min at 4 °C, and discard the supernatant.

24. Wash the beads for 5 min with 1 ml of each of the following buffers and centrifuge at 3,800 g for 2 min at 4 °C between the washes: once with low salt wash buffer, once with high salt wash buffer, once with LiCl wash buffer and twice with TE buffer. CRITICAL STEP All wash buffer AND TE buffer need to be cold before use.

25. Elute the protein-DNA complexes from the beads by adding 240 µl elution buffer and incubating at room temperature for 15 min on a rotating wheel.

26. Centrifuge at 3,800*g* for 2 min at 4 °C, and transfer the supernatant (~230 µl) to a new 1.7 ml microcentrifuge tube.

27. Repeat this step by adding 190 µl elution buffer to the beads and incubating at room temperature for 30 min on a rotating wheel. Centrifuge at 3,800*g* for 2 min at 4 °C, and combine the supernatant (~180 µl) with the previous eluate (total volume will be 410 µl for each of ChIP and mock). For input control, add 360 µl elution buffer to the 50 µl input chromatin sample to reach the same volume as the others.

Reverse crosslinking and protein digestion TIMING overnight and 1 h 30 min

28. Reverse crosslink by adding 16.4 µl of 5 M NaCl to 410 µl of each tube, mixing gently and incubating at 65 °C for at least 6 h to overnight.

29. Add 8.5 µl 0.5 M EDTA, 17 µl 1 M Tris-HCl pH 7.5, 1.7 µl proteinase K (20 mg/ml) and 4.3 µl 1 M CaCl2 to each tube. Incubate at 45 °C for 1.5 h for protein digestion.

DNA precipitation TIMING 2 h 30 min

30. Add equal volume (450 µl) of phenol/chloroform/isoamyl alcohol to each tube and vortex gently.

31. Centrifuge each samples in a microcentrifuge at 13,800*g* for 15 min at 4 °C and transfer the supernatant (~410 µl) to a new 1.7 ml microcentrifuge tube.

32. Add 2.5 volume (1.025 ml) of 100% ethanol, 1/10 volume (41 µl) of 3 M NaoAc (pH 5.2) and 4.1 µl glycogen (20 mg/ml) to each tube. Incubate the tubes for 2 h at -80 °C to precipitate the DNA.

33. Centrifuge each samples in a microcentrifuge at 13,800*g* for 15 min at 4 °C

34. Discard the supernatant, add 700 µl of 70% ethanol to the pellet and vortex shortly.

35. Centrifuge each sample in a microcentrifuge at 13,800*g* for 7 min at 4 °C.

36. Discard the supernatant and dry the pellet at room temperature.

37. Dissolve the DNA in 51 µl of TE (pH 7.5).

38. Measure the DNA concentration using a NanoDrop. TROUBLESHOOTING

Preparation of samples for qPCR and ChIP-seq library construction TIMING 2 h 30 min

39. Aliquot 30 µl of mock or ChIP to new PCR tubes and dilute the DNA by adding 30 µl of TE (pH 7.5) (2X dilution) for qPCR analysis. Save the remaining ~20 µl of ChIP DNA stock for ChIP-seq library construction.

40. Aliquot 12 µl of input to a new PCR tube and dilute the DNA by adding 48 µl of TE (pH 7.5) (5X dilution) for qPCR. 5X dilute the remaining input for ChIP-seq library construction (e.g. dilute 30 µl of input by adding 120 µl of TE (pH 7.5)).

41. Make sure that you have samples for qPCR analysis (2X-diluted mock and ChIP, and 5X-diluted input) and for ChIP-seq library construction (ChIP and mock stock, and 5X-diluted input). PAUSE POINT The DNA can be stored at -80 °C for 3 months without significant loss of DNA quality. CRITICAL STEP Carry out qPCR to test for DNA enrichment in ChIP, mock (negative control) and input (positive control) samples, before preparing ChIP-seq libraries in next steps. The ChIP-seq library construction requires >10 ng of qPCR verified or input DNA with my suggestion of ~50 ng. TROUBLESHOOTING

End Repair of ChIP DNA TIMING 40 min

42. Dilute DNA polymerase I, Large (Klenow) Fragment by mixing 1 µl of enzyme with 4 µl of sterile water in a new 1.7 ml microcentrifuge tube.

43. Combine and mix the following components in a new 1.7 ml low-adhesion microcentrifuge tube, and incubate at 20 °C for 30 min.

ChIP DNA	8	µl
Phosphorylation Reaction Buffer (10X)	5	µl
dNTP mix (10 mM each)	2	µl
T4 DNA polymerase	1	µl
Diluted DNA Pol I Klenow Fragment	1	µl
T4 Polynucleotide Kinase	1	µl
ddH_2O	32	µl
Total volume	50	µl

Purifying DNA: clean up the End Repaired DNA using AMPure XP Beads TIMING 40 min

44. Vortex AMPure XP Beads to resuspend before use.

45. Add 90 µl (1.8X) of resuspended AMPure XP Beads to the reaction (50 µl), vortex shortly (3 sec) and incubate for 5 min at room temperature.

46. Put the tube at the Magnetic particle concentrator and wait until the beads are completely separated from the supernatant (~5 min).

47. Carefully remove and discard the supernatant. CRITICAL STEP Do not disturb the breads which attach the DNA.

48. Add 200 µl of 80% ethanol to the tube, which is still at the Magnetic particle concentrator, and incubate at room temperature for 30 sec.

49. Carefully remove and discard the supernatant. CRITICAL STEP Do not disturb the breads which attach the DNA.

50. Repeat Step 48-49 once.

51. Dry breads in the tube, which is still at the Magnetic particle concentrator, with the lid open at room temperature for 10 min and then separate the tube from the Magnetic particle concentrator.

52. Add 36 µl of 0.1X TE to the tube to elute DNA from beads and mix well by pipetting up and down.

53. Put the tube at the Magnetic particle concentrator and wait until the beads are completely separated from the supernatant (~5 min).

54. Transfer 34 µl of the supernatant to a new 1.7 ml low-adhesion microcentrifuge tube. PAUSE POINT The DNA can be stored at -20 °C until use.

dA-tailing of End Repaired DNA TIMING 40 min

55. Combine and mix the following components in the tube, which contains 34 µl of end repaired DNA, and incubate at 37 °C for 30 min.

End Repaired DNA	34	µl
NEBuffer 2	5	µl
dATP (1.0 mM)	10	µl
Klenow Fragment (3′→ 5′ exo–)	1	µl
Total volume	50	µl

Purifying DNA: clean up the End Repaired DNA using AMPure XP Beads TIMING 40 min

56. Repeat Step 44-51.

57. Add 12 µl of 0.1X TE to the tube to elute DNA from beads and mix well by pipetting up and down.

58. Put the tube at the Magnetic particle concentrator and wait until the beads are completely separated from the supernatant (~5 min).

59. Transfer 10 µl of the supernatant to a new 1.7 ml low-adhesion microcentrifuge tube. PAUSE POINT The DNA can be stored at -20 °C until use.

Adaptor ligation of dA-tailed DNA TIMING 30 min

60. Combine and mix the following components in the tube, which contains 10 µl of dA-tailed DNA, and incubate at 20 °C for 15 min.

End Repaired, dA-tailed DNA	10	µl
Quick Ligation Reaction Buffer (2X)	15	µl
Diluted NEBNext adaptor (1.5 µM)	1	µl
Quick T4 DNA Ligase	1	µl
ddH_2O	3	µl
Total volume	30	µl

Purifying DNA: clean up the End Repaired DNA using AMPure XP Beads TIMING

40 min

61. Vortex AMPure XP Beads to resuspend before use.

62. Add 54 µl (1.8X) of resuspended AMPure XP Beads to the reaction (30 µl), vortex shortly (3 sec) and incubate for 5 min at room temperature.

63. Repeat Step 46-51.

64. Add 103 µl of 0.1X TE to the tube to elute DNA from beads and mix well by pipetting up and down.

65. Put the tube at the Magnetic particle concentrator and wait until the beads are completely separated from the supernatant (~5 min).

66. Transfer 100 µl of the supernatant to a new 1.7 ml low-adhesion microcentrifuge tube. PAUSE POINT The DNA can be stored at -20 °C until use.

Size selection of adaptor-ligated DNA TIMING 45 min

67. Vortex AMPure XP Beads to resuspend before use.

68. Add 50 µl (0.5X) of resuspended AMPure XP Beads to the reaction (100 µl), vortex shortly (3 sec) and incubate for 5 min at room temperature.

69. Put the tube at the Magnetic particle concentrator and wait until the beads are completely separated from the supernatant (~5 min).

70. Carefully transfer the supernatant to a new 1.7 ml low-adhesion microcentrifuge tube. CRITICAL STEP Do not discard the supernatant. Discard beads that contain the large fragments.

71. Add 30 µl (0.3X) of resuspended AMPure XP Beads to the reaction (100 µl), vortex shortly (3 sec) and incubate for 5 min at room temperature.

72. Carefully remove and discard the supernatant. CRITICAL STEP Do not disturb the breads which attach the DNA.

73. Add 200 µl of 80% ethanol to the tube, which is still at the Magnetic particle concentrator, and incubate at room temperature for 30 sec.

74. Carefully remove and discard the supernatant. CRITICAL STEP Do not disturb the breads which attach the DNA.

75. Repeat Step 73-74 once.

76. Dry breads in the tube, which is still at the Magnetic particle concentrator, with the lid open at room temperature for 10 min and then separate the tube from the Magnetic particle concentrator.

77. Add 25 µl of 0.1X TE to the tube to elute DNA from beads and mix well by pipetting up and down.

78. Put the tube at the Magnetic particle concentrator and wait until the beads are completely separated from the supernatant (~5 min).

79. Transfer 23 µl of the supernatant to a new PCR tube. CRITICAL STEP Be sure not to transfer any beads.

PCR enrichment of adaptor-ligated DNA TIMING 2 h

80. Combine and mix the following components in a new PCR tube. CRITICAL STEP Multiplexing allows sequencing of multiple samples as a pool in a lane, which is important for cost effectiveness. For multiplexing, unique index primer needs to be added in the mixture.

Adaptor ligated DNA	23	µl
NEBNext High-Fidelity 2X PCR Master Mix	25	µl

Universal PCR Primer (12.5 µM)	1	µl
Index Primer (12.5 µM)	1	µl
	50	µl

81. Perform PCR using the following conditions.

Cycle step	Temp (°C)	Time	Cycles
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	
Annealing	65	30 sec	18
Extension	72	30 sec	
Final extension	72	5 min	1
	4	hold	

Purifying DNA: clean up PCR products TIMING 40 min

82. Vortex AMPure XP Beads to resuspend before use.

83. Add 50 µl (1X) of resuspended AMPure XP Beads to the reaction (50 µl), vortex shortly (3 sec) and incubate for 5 min at room temperature.

84. Repeat Step 46-51.

85. Add 22 µl of 0.1X TE to the tube to elute DNA from beads and mix well by pipetting up and down.

86. Put the tube at the Magnetic particle concentrator and wait until the beads are completely separated from the supernatant (~5 min).

87. Transfer 20 µl of the supernatant to a new PCR tube. PAUSE POINT The DNA can be stored at -20 °C until use.

88. Measure the DNA concentration using NanoDrop. TROUBLESHOOTING

Gel running for PCR verification of TIMING 45 min

89. Combine and mix the following components in a new PCR tube: 5 µl of ChIP-seq library, 5 µl of ddH_2O and 2 µl of 40% sucrose.

90. Run the mixture on 1.5% agarose gel.

91. (Optional) For further checking size distribution, perform electropherogram analysis using an Agilent Bioanalyzer (Figure 5.2A).

5.4 Timing

Day 1

Steps 1-6, tissue collection and crosslinking: 1 h

Steps 7-17, nuclear isolation and chromatin fragmentation: 4 h

Steps 18-21, immunoprecipitation: 1 h 30 min and overnight

Day 2

Steps 22-27, washing and elution: 4 h

Step 28, reverse crosslinking: 30 min and overnight

Day 3

Step 29, protein digestion: 1 h 30 min

Steps 30-38, DNA precipitation: 2 h 30 min

Steps 39-41, sample preparation for qPCR analysis and ChIP-seq library construction: 1 h

Day 4

Steps 42-66, end repair, dA-tailing and adaptor ligation: 4 h

Steps 67-79, size selection of adaptor-ligated DNA: 45 min

Steps 80-88, PCR enrichment: 2 h

Steps 89-90, Gel running, 45 min

5.5. Troubleshooting

Troubleshooting advice can be found in Table 5.1.

5.6 Anticipated results

I have here developed a robust, time-efficient and low-cost protocol for a combination of ChIP and ChIP-seq library construction in maize tissue. From my experience, yielding high-quality ChIP DNA and ChIP-seq library depends on many different steps in this protocol. For example, an efficient sonication can be a good indicator of that; with the optimized condition for the sonication, the sample from Step 16 should have a smear from 0.2 to 2-kb, but concentrated ~500-bp compared to one from Step 13. A chromatin yield of ~30 μg of DNA per g of fresh tissue should be expected to give a good DNA enrichment. By using this protocol and a native antibody against a TF, >10-fold enrichment of bound genomic DNA relative to one in mock sample will be achieved as I did in my study (See Chapter 4) [Ko *et al*., in revision]. In the library construction, to avoid PCR artifacts (e.g. duplicates) and possible false negatives, PCR cycles should be minimized as 18 cycles are suggested in this protocol. In case of using a native antibody against TFs, a large portion of mapped reads is generally expected to be PCR duplicates, depending on the specificity of the antibody and PCR cycles (Step 81), and should be removed using appropriate algorithms [136] in following computational works. The size selection (Steps 67-79) of ChIP-seq library construction should yield

fragments from 200 to 800 bp, but concentrated ~400-bp (Figure 5.2A). Finally, high-quality library will be expected to have final concentration >15 ng/μl in 50 μl (>750 ng) at Step 88, which is a plenty of amount for Illumina sequencing platforms. The resulting sequencing data will be subsequently analyzed in the standard ChIP-seq pipeline [194] and expected to show the strong enrichment (Figure 5.2B).

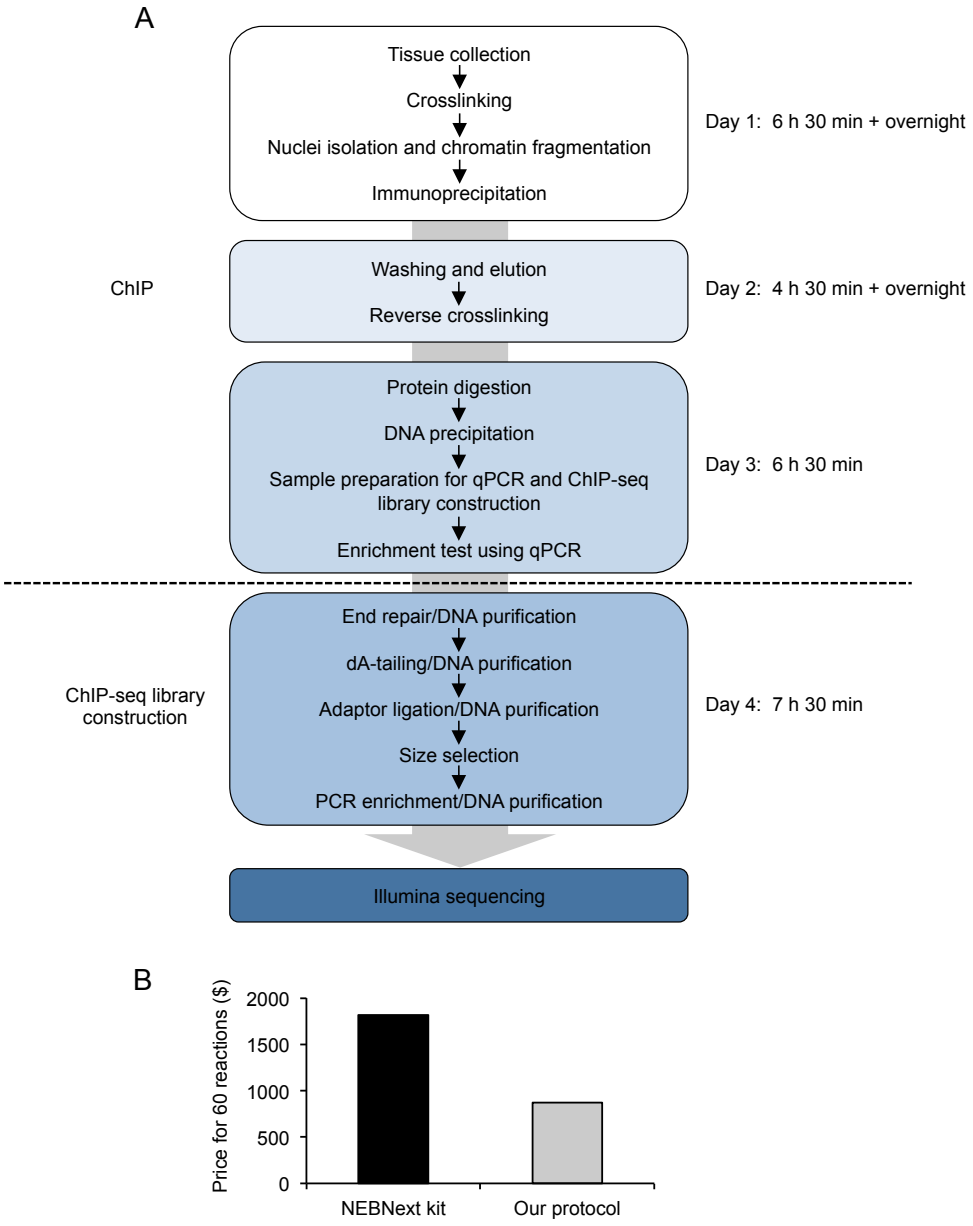

Figure 5.1 Overview of the protocol described in this Chapter.

(A) Workflow and timeline of ChIP and ChIP-seq library construction described in this protocol. The different steps are represented with their expected timelines. Expected timeline was designed to handle 8 samples each for ChIP and ChIP-seq library construction. (B) Comparison of estimated cost using a commercial kit with one using this protocol for ChIP-seq library construction. Black and grey bars represent NEBNext kit (E6200L) and this protocol, respectively. The Y-axis indicates the estimated price, which includes all reagents required for ChIP-seq library construction except for multiplex oligomers and AMPure XP beads.

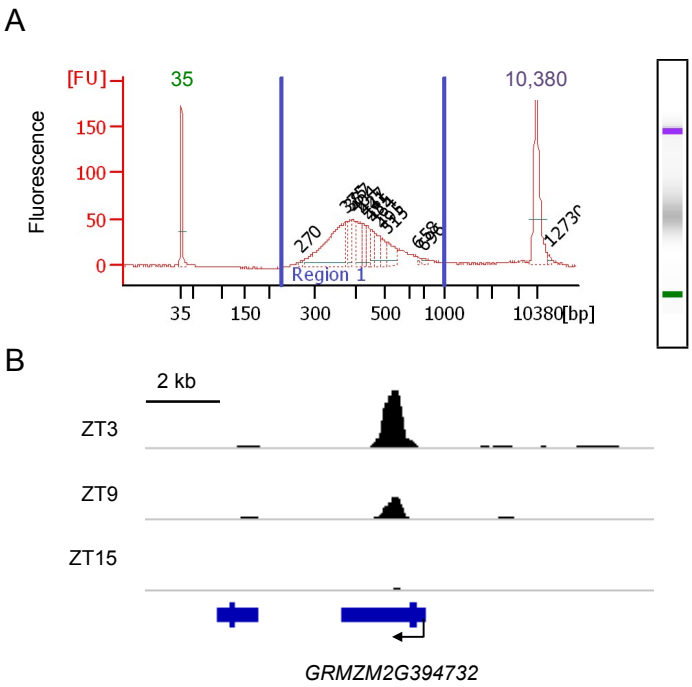

Figure 5.2 Examples of ChIP-seq library construction and sequencing results.

(A) An example of electropherogram analysis in ChIP-seq library resulting from this protocol. The DNA fragment size distribution of a ChIP-seq library is visualized using an Agilent Bioanalyzer. The size distribution is between approximately 300 and 600-bp with a peak intensity at 370-bp. Peaks at 35 and 10,385-bp represent lower and upper molecular markers, respectively. (B) A snapshot of ChIP-seq profiles for ZmCCA1 in 5-d old maize seedlings at different times of a day. The Y-axis indicates duplicate-filtered input subtracted ChIP signals on a same-scale for all time-points. Arrows indicate gene orientation. The screenshot is from Integrative Genomic Viewer (version 2.3.46).

Step	Problem	Possible reason	Solution
3	No visible sign for translucent tissue	Vacuum system does not work properly.	Make use that your vacuum system works well.
		Tissues are not completely submerged.	Shake the flask gently to make sure that tissues are completely submerged when vacuum applies.
		There is too much amount of tissues in the buffer.	Make sure that there is appropriate amount of tissues in the buffer.
17	Low chromatin yield	Not enough amount of starting tissue	Harvest more tissue.
		Cells are not released.	Grind the tissue to a fine powder (Step 7). Make sure that the ground tissue is completely homogenized with Nuclei isolation buffer A.
	Chromatin is not well sonicated.	The sonication does not work well.	Make sure to have the correct condition for the sonication (Step 14). Carefully read the manual of the sonication system, which provides a checklist before running.
		Proteins bound DNA affect the DNA running.	Sometimes, proteins bound to DNA may affect the DNA running in the gel. In case of this, I recommend running the gel after reverse crosslinking (Step 28).
38	Low yield of ChIP DNA	The efficiency of immunoprecipitation is low.	Use high affinity purified antibody.
41	No qPCR signal	qPCR does not work well.	Make sure to have positive (e.g. input) and negative (e.g. mock) controls for the reaction.
		qPCR primers are incorrectly designed.	Make sure to have correct sequences of qPCR primers. I recommend having qPCR primers, of which amplification was already validated in previous studies if possible. In case of designing new primers, use primer software available in your laboratory or online.
	Lower enrichment of ChIP DNA than expected.	The antibody is not specific to the protein of interest.	Use highly purified and specific antibody to the protein of interest.
		The chromatin is not sufficiently sheared ranging from 0.2 to 2 kb.	Make sure to have a good sonication result (Step 17).
		Preclearing or washing does not sufficiently remove nonspecific binding or insoluble material, respectively.	Increase pre-clearing time to 1.5 h (Step 19). Adjust the stringency of the washing solutions and the number of washes.
88	Concentration of the final library is lower than expected.	Loss of DNA during purification and size-selection steps	Loss of DNA frequently occurs during DNA purification (Steps 44-54, 56-59, 61-66 and 82-88) and size selection (Steps 67-79) steps. Make sure to carefully follow all procedures in the steps.
		Incomplete adaptor ligation reduces effectiveness of PCR.	Make sure to have the correct incubation condition and components.

Table 5.1 Troubleshooting table

APPENDIX. Summary of ZmCCA1s ChIP-seq mapping

Genotype	ZT	Sample type	Biol. rep.	# Raw read-pairs	% Mapping rate	# Uniquely mapped* read-pairs (with duplicates)	# Uniquely mapped* read-pairs (without duplicates)	# Down-sampled uniquely mapped reads (without duplicates)
B73	3	ChIP	1	18,812,717	63.46	5,293,149	2,135,504	2,096,442
BM	3	ChIP	1	19,241,776	55.04	4,862,904	1,152,116	2,096,146
MB	3	ChIP	1	19,391,273	57.58	4,980,079	1,170,503	2,095,434
Mo17	3	ChIP	1	21,937,289	61.83	5,601,530	2,015,033	2,097,538
B73	3	Input	1	33,120,634	71.35	9,440,228	9,129,653	10,281,272
BM	3	Input	1	39,830,759	72.53	11,094,039	10,493,882	10,284,930
MB	3	Input	1	31,056,428	72.04	8,572,234	8,242,457	10,279,580
Mo17	3	Input	1	38,226,163	71.24	9,761,317	9,662,472	10,273,896
B73	3	ChIP	2	17,730,331	56.06	4,556,943	1,670,482	2,094,662
BM	3	ChIP	2	12,721,100	54.30	3,083,729	1,144,601	2,095,618
MB	3	ChIP	2	9,912,528	55.07	2,308,541	1,047,850	2,095,688
Mo17	3	ChIP	2	16,002,796	62.29	4,194,671	1,838,532	2,094,916
B73	3	Input	2	28,044,403	70.72	7,921,851	7,567,949	10,284,620
BM	3	Input	2	40,580,485	71.00	11,078,249	10,297,926	10,279,786
MB	3	Input	2	23,677,951	70.50	6,377,428	6,162,894	10,280,814
Mo17	3	Input	2	25,817,966	70.28	6,547,885	6,426,413	10,281,472
B73	9	ChIP	1	15,804,920	55.80	3,841,784	2,411,046	2,095,688
BM	9	ChIP	1	20,531,767	61.49	5,331,011	3,129,300	2,097,152
MB	9	ChIP	1	14,508,856	59.43	3,730,487	2,109,493	2,096,694
Mo17	9	ChIP	1	16,072,514	62.01	4,126,349	2,537,406	2,094,548
B73	9	Input	1	45,698,716	70.80	12,872,822	12,495,611	10,291,344
BM	9	Input	1	45,525,047	72.13	12,562,180	12,198,961	10,293,288
MB	9	Input	1	32,734,396	71.88	9,076,047	8,776,820	10,283,360
Mo17	9	Input	1	22,502,176	70.02	5,730,267	5,600,457	10,282,640
B73	9	ChIP	2	15,976,189	59.28	4,189,399	2,441,427	2,094,674
BM	9	ChIP	2	16,855,110	63.10	4,587,095	2,812,749	2,096,910
MB	9	ChIP	2	12,028,647	66.05	3,351,021	2,310,537	2,096,344
Mo17	9	ChIP	2	10,590,660	57.10	2,554,970	1,354,164	2,095,864
B73	9	Input	2	33,023,886	71.46	9,325,995	9,116,463	10,286,712
BM	9	Input	2	32,274,256	71.61	8,951,030	8,628,640	10,285,338
MB	9	Input	2	28,651,362	71.90	7,913,713	7,643,434	10,284,158
Mo17	9	Input	2	23,882,879	70.22	6,005,211	5,877,309	10,287,290
B73	15	ChIP	1	20,288,627	61.04	5,533,406	2,943,357	2,096,714
BM	15	ChIP	1	18,934,373	62.00	5,008,290	2,794,833	2,094,138

MB	15	ChIP	1	16,098,005	60.58	4,112,479	1,787,244	2,093,534
Mo17	15	ChIP	1	13,538,518	48.88	2,661,517	1,772,091	2,096,864
B73	15	Input	1	26,295,719	71.00	7,443,824	7,113,794	10,281,406
BM	15	Input	1	41,621,018	71.67	11,300,917	10,795,524	10,282,566
MB	15	Input	1	33,011,407	71.82	9,039,475	8,832,426	10,285,672
Mo17	15	Input	1	21,580,961	69.55	5,305,022	5,141,882	10,283,752
B73	15	ChIP	2	21,885,820	62.58	5,867,350	2,930,744	2,096,026
BM	15	ChIP	2	25,624,124	69.35	6,859,825	5,219,660	2,096,520
MB	15	ChIP	2	17,109,424	67.06	4,578,249	3,292,002	2,097,478
Mo17	15	ChIP	2	14,911,933	62.09	3,700,185	2,054,424	2,096,774
B73	15	Input	2	28,337,553	71.46	7,865,577	7,590,052	10,286,728
BM	15	Input	2	39,869,030	72.33	10,849,595	10,545,382	10,281,684
MB	15	Input	2	32,179,907	72.69	8,745,697	8,487,720	10,279,704
Mo17	15	Input	2	37,958,599	71.16	9,569,442	9,242,678	10,285,712

*“Uniquely mapped reads” means concordantly paired-reads mapped to the genome exactly 1 time in this study.

References

1. Duvick, D.N., *Biotechnology in the 1930s: the development of hybrid maize.* Nat Rev Genet, 2001. **2**(1): p. 69-74.
2. Chen, Z.J., *Molecular mechanisms of polyploidy and hybrid vigor.* Trends Plant Sci, 2010. **15**(2): p. 57-71.
3. Birchler, J.A., et al., *Heterosis.* Plant Cell, 2010. **22**(7): p. 2105-2112.
4. Chen, Z.J. and J.A. Birchler, *Polyploid and hybrid genomics*. 2013: New York: Wiley-Blackwell.
5. Steinmetz, L.M., et al., *Dissecting the architecture of a quantitative trait locus in yeast.* Nature, 2002. **416**(6878): p. 326-330.
6. Tirosh, I., et al., *A yeast hybrid provides insight into the evolution of gene expression regulation.* Science, 2009. **324**(5927): p. 659-662.
7. Hedgecock, D., D.J. McGoldrick, and B.B. L., *Hybrid vigor in Pacific oysters: an experimental approach using crosses among inbred lines.* Aquaculture, 1995. **137**(1): p. 285-298.
8. Cassady, J.P., L.D. Young, and K.A. Leymaster, *Heterosis and recombination effects on pig growth and carcass traits.* J Anim Sci, 2002. **80**(9): p. 2286-2302.
9. Sagebiel, J.A., et al., *Effect of heterosis and maternal influence on gestation length and birth weight in reciprocal crosses among Angus, Charolais and Hereford cattle.* J Anim Sci, 1973. **37**(6): p. 1273-1278.
10. Ishikawa, A., *Mapping an overdominant quantitative trait locus for heterosis of body weight in mice.* J Hered, 2009. **100**(4): p. 501-504.
11. Gemmell, N.J. and J. Slate, *Heterozygote advantage for fecundity.* PLoS One, 2006. **1**: p. e125.
12. Birchler, J.A., D.L. Auger, and N.C. Riddle, *In search of the molecular basis of heterosis.* Plant Cell, 2003. **15**(10): p. 2236-9.
13. Crow, J.F., *90 years ago: the beginning of hybrid maize.* Genetics, 1998. **148**(3): p. 923-928.
14. Darwin, C.R., *The effects of cross and self fertilisation in the vegetable kingdom*. 1876: John Murray.
15. Pusey, A. and M. Wolf, *Inbreeding avoidance in animals.* Trends Ecol Evol, 1996. **11**(5): p. 201-206.
16. Shull, G.H., *The composition of a field of maize.* Am Breeders Assoc Rep, 1908. **4**: p. 296-301.
17. East, E.M., *Heterosis.* Genetics, 1936. **21**(4): p. 375-397.
18. Moll, R.H., et al., *The Relationship of Heterosis and Genetic Divergence in Maize.* Genetics, 1965. **52**(1): p. 139-44.
19. Karpechenko, G.D., *Polyploid Hybrids of Raphanus sativus L. x Brassica oleracea L.* Bull. Appl. Bot., 1927. **17**: p. 305-410.

20. Eshed, Y. and D. Zamir, *An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL.* Genetics, 1995. **141**(3): p. 1147-1162.
21. Jones, D.F., *The effect of inbreeding and crossbreeding upon development.* Proc Natl Acad Sci USA, 1918. **4**(8): p. 246-250.
22. *FAOSTAT. Food and Agriculture (Organization of the United Nations)*. 2015, Food and Agriculture Organization of the Unites Nations: http://faostat3.fao.org/home/E.
23. Russell, W.A., *Comparative performance for maize hybrids representing different eras of maize breeding*, in *Proc 29th Annu Corn Sorghum Res Conf*. 1974. p. 81-101.
24. Lippman, Z.B. and D. Zamir, *Heterosis: revisiting the magic.* Trends Genet, 2007. **23**(2): p. 60-66.
25. Goff, S.A. and Q. Zhang, *Heterosis in elite hybrid rice: speculation on the genetic and biochemical mechanisms.* Curr Opin Plant Biol, 2013. **16**(2): p. 221-227.
26. Duvick, D.N., *Heterosis: feeding people and protecting natural resources*, in *The genetics and exploitation of heterosis in crops*. 1999, American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. p. 19-29.
27. Godfray, H.C., et al., *Food security: the challenge of feeding 9 billion people.* Science, 2010. **327**(5967): p. 812-818.
28. Foley, J.A., et al., *Solutions for a cultivated planet.* Nature, 2011. **478**(7369): p. 337-342.
29. Ray, D.K., et al., *Yield Trends Are Insufficient to Double Global Crop Production by 2050.* PLoS One, 2013. **8**(6): p. e66428.
30. Birchler, J.A., D.L. Auger, and N.C. Riddle, *In search of the molecular basis of heterosis.* Plant Cell, 2003. **15**(10): p. 2236-2239.
31. Chen, Z.J., *Genomic and epigenetic insights into the molecular bases of heterosis.* Nat Rev Genet, 2013. **14**(7): p. 471-482.
32. Schnable, P.S. and N.M. Springer, *Progress toward understanding heterosis in crop plants.* Annu Rev Plant Biol, 2013. **64**: p. 71-88.
33. Li, L., et al., *Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids.* Genetics, 2008. **180**(3): p. 1725-42.
34. Swanson-Wagner, R.A., et al., *All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents.* Proc Natl Acad Sci USA, 2006. **103**(18): p. 6805-6810.
35. Ni, Z., et al., *Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids.* Nature, 2009. **457**(7227): p. 327-331.
36. Ng, D.W., et al., *A Role for CHH Methylation in the Parent-of-Origin Effect on Altered Circadian Rhythms and Biomass Heterosis in Arabidopsis Intraspecific Hybrids.* Plant Cell, 2014. **26**(6): p. 2430-2440.
37. Miller, M., C. Zhang, and Z.J. Chen, *Ploidy and Hybridity Effects on Growth Vigor and Gene Expression in Arabidopsis thaliana Hybrids and Their Parents.* G3 (Bethesda), 2012. **2**(4): p. 505-13.

38. Edgar, R.S., et al., *Peroxiredoxins are conserved markers of circadian rhythms.* Nature, 2012. **485**(7399): p. 459-464.
39. de Mairan, J., *Observation botanique.* Hist Acad Roy Sci, 1729: p. 35-36.
40. Greenham, K. and C.R. McClung, *Integrating circadian dynamics with physiological processes in plants.* Nat Rev Genet, 2015. **16**(10): p. 598-610.
41. Bell-Pedersen, D., et al., *Circadian rhythms from multiple oscillators: lessons from diverse organisms.* Nat Rev Genet, 2005. **6**(7): p. 544-556.
42. Doherty, C.J. and S.A. Kay, *Circadian control of global gene expression patterns.* Annu Rev Genet, 2010. **44**(1): p. 419-444.
43. Harmer, S.L., *The circadian system in higher plants.* Annu Rev Plant Biol, 2009. **60**: p. 357-377.
44. Dunlap, J.C., *Molecular bases for circadian clocks.* Cell, 1999. **96**(2): p. 271-290.
45. McClung, C.R., *Plant circadian rhythms.* Plant Cell, 2006. **18**(4): p. 792-803.
46. Más, P., *Circadian clock function in Arabidopsis thaliana: time beyond transcription.* Trends Cell Biol, 2008. **18**(6): p. 273-281.
47. Pittendrigh, C.S., *On temperature independence in the clock system controlling emergence time in Drosophila.* Proc Natl Acad Sci USA, 1954. **40**(10): p. 1018-1029.
48. Bass, J. and J.S. Takahashi, *Circadian integration of metabolism and energetics.* Science, 2010. **330**(6009): p. 1349-1354.
49. Pruneda-Paz, J.L. and S.A. Kay, *An expanding universe of circadian networks in higher plants.* Trends Plant Sci, 2010. **15**(5): p. 259-265.
50. Dodd, A.N., et al., *Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage.* Science, 2005. **309**(5734): p. 630-633.
51. Green, R.M., et al., *Circadian rhythms confer a higher level of fitness to Arabidopsis plants.* Plant Physiol, 2002. **129**(2): p. 576-584.
52. Graf, A., et al., *Circadian control of carbohydrate availability for growth in Arabidopsis plants at night.* Proc Natl Acad Sci USA, 2010. **107**(20): p. 9458-9463.
53. McClung, C.R., *Wheels within wheels: new transcriptional feedback loops in the Arabidopsis circadian clock.* F1000Prime Rep, 2014. **6**.
54. Khan, S., S.C. Rowe, and F.G. Harmon, *Coordination of the maize transcriptome by a conserved circadian clock.* BMC Plant Biol, 2010. **10**(1): p. 126.
55. Evert, R.F., W.A. Russin, and A.M. Bosabalidis, *Anatomical and ultrastructural changes associated with sink-to-source transition in developing maize leaves.* Int J Plant Sci, 1996. **157**(3): p. 247-261.
56. Kalt-Torres, W., et al., *Diurnal changes in maize leaf photosynthesis : I. Carbon exchange rate, assimilate export rate, and enzyme activities.* Plant Physiol, 1987. **83**(2): p. 283-288.
57. Sadok, W., et al., *Leaf growth rate per unit thermal time follows QTL-dependent daily patterns in hundreds of maize lines under naturally fluctuating conditions.* Plant Cell Environ, 2007. **30**(2): p. 135-46.

58. Hayes, K.R., et al., *Maize global transcriptomics reveals pervasive leaf diurnal rhythms but rhythms in developing ears are largely limited to the core oscillator.* PLoS One, 2010. **5**(9): p. e12887.
59. Crow, J.F., *Mid-century controversies in population genetics.* Annu Rev Genet, 2008. **42**: p. 1-16.
60. Murphy, R.L., et al., *Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum.* Proc Natl Acad Sci USA, 2011. **108**(39): p. 16469-16474.
61. Bendix, C., et al., *The circadian clock-associated gene gigantea1 affects maize developmental transitions.* Plant Cell Environ, 2013. **36**(7): p. 1-12.
62. Song, G.S., et al., *Comparative transcriptional profiling and preliminary study on heterosis mechanism of super-hybrid rice.* Mol Plant, 2010. **3**(6): p. 1012-25.
63. Watanabe, S., et al., *A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering.* Genetics, 2011. **188**(2): p. 395-407.
64. Yang, S., et al., *CONSTANS is a photoperiod regulated activator of flowering in sorghum.* BMC Plant Biol, 2014. **14**: p. 148.
65. Müller, N.A., et al., *Domestication selected for deceleration of the circadian clock in cultivated tomato.* Nat Genet, 2016. **48**(1): p. 89-93.
66. Buckler, E.S., et al., *The genetic architecture of maize flowering time.* Science, 2009. **325**(5941): p. 714-718.
67. Chen, C., et al., *PICARA, an analytical pipeline providing probabilistic inference about a priori candidates genes underlying genome-wide association QTL in plants.* PLoS One, 2012. **7**(11): p. e46596.
68. Cheng, S.H., et al., *Progress in research and development on hybrid rice: a super-domesticate in China.* Ann Bot, 2007. **100**(5): p. 959-966.
69. Yuan, L.P., *Hybrid rice breeding in China*, in *Advances in Hybrid Rice Technology*. 1998, Philippines: International Rice Research Institute. p. 27-33.
70. Lin, S. and L. Yuan, *Hybrid rice breeding in China*, in *In Innovative approaches to rice breeding*. 1980, International Rice Research Institute, Los Banos, Philippines. p. 35-51.
71. Longin, C.F.H., et al., *Hybrid breeding in autogamous cereals.* Theor Appl Genet, 2012. **125**(6): p. 1087-1096.
72. Wendel, J.F. and R.C. Cronn, *Polyploidy and the evolutionary history of cotton.* Adv Agron, 2003. **78**: p. 139-186.
73. Guan, X., Q. Song, and Z.J. Chen, *Polyploidy and small RNA regulation of cotton fiber development.* Trends Plant Sci, 2014. **19**(8): p. 516-528.
74. Dubcovsky, J. and J. Dvorak, *Genome plasticity a key factor in the success of polyploid wheat under domestication.* Science, 2007. **316**(5833): p. 1862-1866.
75. Huang, S., et al., *Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat.* Proc Natl Acad Sci USA, 2002. **99**(12): p. 8133-8138.
76. Feldman, M., et al., *Genomic asymmetry in allopolyploid plants: wheat as a model.* J Exp Bot, 2012. **63**(14): p. 5045-5059.

77. Zanoni, U. and J.W. Dudley, *Comparison of different methods of identifying inbreds useful for improving elite maize hybrids.* Crop Sci, 1989. **29**(3): p. 577-582.
78. Auger, D.L., E.M. Peters, and J.A. Birchler, *A genetic test of bioactive gibberellins as regulators of heterosis in maize.* J Hered, 2005. **96**(5): p. 614-617.
79. Flint-Garcia, S.A., et al., *Heterosis is prevalent for multiple traits in diverse maize germplasm.* PLoS One, 2009. **4**(10): p. e7433.
80. Meyer, R.C., et al., *Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids.* Plant J, 2012. **71**(4): p. 669-683.
81. Fujimoto, R., et al., *Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity.* Proc Natl Acad Sci USA, 2012. **109**(18): p. 7109-7114.
82. Hoecker, N., et al., *Manifestation of heterosis during early maize (Zea mays L.) root development.* Theor Appl Genet, 2006. **112**(3): p. 421-429.
83. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-675.
84. Belo, A., et al., *Allelic genome structural variations in maize detected by array comparative genome hybridization.* Theor Appl Genet, 2010. **120**(2): p. 355-367.
85. Pick, T.R., et al., *Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation.* Plant Cell, 2011. **23**(12): p. 4208-4220.
86. Mehta, H. and K.R. Sarkar, *Heterosis for leaf photosynthesis, grain yield and yield components in maize.* Euphytica, 1992. **61**(2): p. 161-168.
87. Sharma, R.K., B. Griffing, and R.L. Scholl, *Variations among races of Arabidopsis thaliana (L.) heynh for survival in limited carbon dioxide.* Theor Appl Genet, 1979. **54**(1): p. 11-15.
88. Wang, L., et al., *Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice.* Nat Biotechnol, 2014. **32**(11): p. 1158-1165.
89. Wang, Z.Y. and E.M. Tobin, *Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression.* Cell, 1998. **93**(7): p. 1207-1217.
90. Sekhon, R.S., et al., *Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays.* PLoS One, 2013. **8**(4): p. e61005.
91. Miller, M., C. Zhang, and Z.J. Chen, *Ploidy and Hybridity Effects on Growth Vigor and Gene Expression in Arabidopsis thaliana Hybrids and Their Parents.* G3, 2012. **2**(4): p. 505-513.
92. Paz, M.M., et al., *Assessment of conditions affecting Agrobacterium-mediated soybean transformation using the cotyledonary node explant.* Euphytica, 2004. **136**(2): p. 167-179.
93. Mann, D.G., et al., *Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (Panicum virgatum L.) and other monocot species.* Plant Biotechnol J, 2012. **10**(2): p. 226-36.

94. Frame, B.R., et al., *Agrobacterium tumefaciens-mediated transformation of maize embryos using a standard binary vector system.* Plant Physiol, 2002. **129**(1): p. 13-22.
95. Osterman, J.C. and E.S. Dennis, *Molecular analysis of the ADH1-Cm allele of maize.* Plant Mol Biol, 1989. **13**(2): p. 203-12.
96. Richardson, A.D., S.P. Duigan, and G.P. Berlyn, *An evaluation of noninvasive methods to estimate foliar chlorophyll content.* New Phytologist, 2002. **153**(1): p. 185-194.
97. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-2948.
98. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol, 1987. **4**(4): p. 406-425.
99. Zuckerkandl, E. and L. Pauling, *Evolutionary divergence and convergence in proteins*, in *Evolving genes and proteins*. 1965, Academic Press: New York. p. 97-166.
100. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0.* Mol Biol Evol, 2013. **30**(12): p. 2725-2729.
101. Clough, S.J. and A.F. Bent, *Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana.* Plant J, 1998. **16**: p. 735-743.
102. Plautz, J.D., et al., *Quantitative analysis of Drosophila period gene transcription in living animals.* J Biol Rhythms, 1997. **12**(3): p. 204-217.
103. Schnable, P.S., et al., *The B73 maize genome: complexity, diversity, and dynamics.* Science, 2009. **326**(5956): p. 1112-1115.
104. Schnable, J.C., N.M. Springer, and M. Freeling, *Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss.* Proc Natl Acad Sci USA, 2011. **108**(10): p. 4069-4074.
105. Seo, P.J., et al., *A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in Arabidopsis.* Plant Cell, 2012. **24**(6): p. 2427-2442.
106. Shen, H., et al., *Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids.* Plant Cell, 2012. **24**(3): p. 875-892.
107. Harmer, S.L. and S.A. Kay, *Positive and negative factors confer phase-specific circadian regulation of transcription in Arabidopsis.* Plant Cell, 2005. **17**(7): p. 1926-1940.
108. Christensen, A.H. and P.H. Quail, *Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants.* Transgenic Res, 1996. **5**(3): p. 213-8.
109. Murakami, M., et al., *Comparative overviews of clock-associated genes of Arabidopsis thaliana and Oryza sativa.* Plant and cell physiology, 2007. **48**(1): p. 110.
110. Filichkin, S.A., et al., *Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules.* PLoS One, 2011. **6**(6): p. e16907.

111. Izawa, T., et al., *Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field.* Plant Cell, 2011. **23**(5): p. 1741-1755.
112. Takata, N., et al., *Molecular phylogeny and expression of poplar circadian clock genes, LHY1 and LHY2.* New Phytol, 2009. **181**(4): p. 808-819.
113. Lou, P., et al., *Preferential retention of circadian clock genes during diploidization following whole genome triplication in Brassica rapa.* Plant Cell, 2012. **24**(6): p. 2415-2426.
114. Marcolino-Gomes, J., et al., *Diurnal oscillations of soybean circadian clock and drought responsive genes.* PLoS One, 2014. **9**(1): p. e86402.
115. Campoli, C., et al., *Expression conservation within the circadian clock of a monocot: natural variation at barley Ppd-H1 affects circadian expression of flowering time genes, but not clock orthologs.* BMC Plant Biol, 2012. **12**: p. 97.
116. Faure, S., et al., *Mutation at the circadian clock gene EARLY MATURITY 8 adapts domesticated barley (Hordeum vulgare) to short growing seasons.* Proc Natl Acad Sci USA, 2012. **109**(21): p. 8328-8333.
117. Wang, X., et al., *Robust expression and association of ZmCCA1 with circadian rhythms in maize.* Plant Cell Rep, 2011. **30**(7): p. 1261-1272.
118. Chaw, S.M., et al., *Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes.* J Mol Evol, 2004. **58**(4): p. 424-441.
119. Wray, G.A., et al., *The evolution of transcriptional regulation in eukaryotes.* Mol Biol Evol, 2003. **20**(9): p. 1377-1419.
120. Johnson, D.S., et al., *Genome-wide mapping of in vivo protein-DNA interactions.* Science, 2007. **316**(5830): p. 1497-1502.
121. Park, P.J., *ChIP-seq: advantages and challenges of a maturing technology.* Nat Rev Genet, 2009. **10**(10): p. 669-680.
122. Robertson, G., et al., *Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing.* Nat Methods, 2007. **4**(8): p. 651-657.
123. Teytelman, L., et al., *Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins.* Proc Natl Acad Sci USA, 2013. **110**(46): p. 18602-18607.
124. Van Nostrand, E.L. and S.K. Kim, *Integrative analysis of C. elegans modENCODE ChIP-seq data sets to infer gene regulatory interactions.* Genome Res, 2013. **23**(6): p. 941-953.
125. Teytelman, L., et al., *Impact of chromatin structures on DNA processing for genomic analyses.* PLoS One, 2009. **4**(8): p. e6700.
126. Furey, T.S., *ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions.* Nat Rev Genet, 2012. **13**(12): p. 840-852.
127. Feng, D., et al., *A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism.* Science, 2011. **331**(6022): p. 1315-1319.
128. Cho, H., et al., *Regulation of circadian behaviour and metabolism by REV-ERB-α and REV-ERB-β.* Nature, 2012. **485**(7396): p. 123-7.

129. Rey, G., et al., *Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver.* PLoS Biol, 2011. **9**(2): p. e1000595.
130. Nagel, D.H., et al., *Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis.* Proc Natl Acad Sci U S A, 2015. **112**(34): p. E4802-10.
131. Kamioka, M., et al., *Direct repression of evening genes by CIRCADIAN CLOCK-ASSOCIATED 1 in Arabidopsis circadian clock.* Plant Cell, 2016.
132. Huang, W., et al., *Mapping the core of the Arabidopsis circadian clock defines the network structure of the oscillator.* Science, 2012. **336**(6077): p. 75-79.
133. Nakamichi, N., et al., *Transcriptional repressor PRR5 directly regulates clock-output pathways.* Proc Natl Acad Sci USA, 2012. **109**(42): p. 17123-17128.
134. Xin, M., et al., *Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development.* Plant Cell, 2013. **25**(9): p. 3212-3227.
135. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-359.
136. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.
137. Koike, N., et al., *Transcriptional architecture and chromatin landscape of the core circadian clock in mammals.* Science, 2012. **338**(6105): p. 349-354.
138. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
139. Nicol, J.W., et al., *The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets.* Bioinformatics, 2009. **25**(20): p. 2730-2731.
140. Ramírez, F., et al., *deepTools: a flexible platform for exploring deep-sequencing data.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W187-W191.
141. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-W208.
142. Du, Z., et al., *agriGO: a GO analysis toolkit for the agricultural community.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W64-70.
143. Wolfe, K.H., et al., *Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data.* Proc Natl Acad Sci U S A, 1989. **86**(16): p. 6201-5.
144. Gupta, S., et al., *Quantifying similarity between motifs.* Genome Biol, 2007. **8**(2): p. R24.
145. Yanagisawa, S. and J. Sheen, *Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression.* Plant Cell, 1998. **10**(1): p. 75-89.
146. Li, P., et al., *The developmental dynamics of the maize leaf transcriptome.* Nat Genet, 2010. **42**(12): p. 1060-1067.
147. Wijnen, H. and M.W. Young, *Interplay of circadian clocks and metabolic rhythms.* Annu Rev Genet, 2006. **40**: p. 409-448.

148. Goff, S.A., *A unifying theory for general multigenic heterosis: energy efficiency, protein metabolism, and implications for molecular breeding.* New Phytol, 2011. **189**(4): p. 923-937.
149. Guo, M., et al., *Cell Number Regulator1 affects plant and organ size in maize: implications for crop yield enhancement and heterosis.* Plant Cell, 2010. **22**(4): p. 1057-1073.
150. Stupar, R.M. and N.M. Springer, *Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid.* Genetics, 2006. **173**(4): p. 2199-210.
151. Swanson-Wagner, R.A., et al., *All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents.* Proc Natl Acad Sci U S A, 2006. **103**(18): p. 6805-10.
152. Riddle, N.C., et al., *Gene expression analysis at the intersection of ploidy and hybridity in maize.* Theor Appl Genet, 2010. **120**(2): p. 341-53.
153. Meyer, R.C., et al., *Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids.* The Plant J., 2012. **71**(4): p. 669-683.
154. Groszmann, M., et al., *Intraspecific Arabidopsis hybrids show different patterns of heterosis despite the close relatedness of the parental genomes.* Plant Physiol, 2014. **166**(1): p. 265-80.
155. Goff, S.A., *A unifying theory for general multigenic heterosis: energy efficiency, protein metabolism, and implications for molecular breeding.* New Phytol, 2011. **189**(4): p. 923-37.
156. Wijnen, H. and M.W. Young, *Interplay of circadian clocks and metabolic rhythms.* Annu Rev Genet, 2006. **40**: p. 409-48.
157. Harmer, S.L., *The circadian system in higher plants.* Annu Rev Plant Biol, 2009. **60**: p. 357-77.
158. Harmer, S.L. and S.A. Kay, *Positive and negative factors confer phase-specific circadian regulation of transcription in Arabidopsis.* Plant Cell, 2005. **17**(7): p. 1926-40.
159. Alabadi, D., et al., *Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock.* Science, 2001. **293**(5531): p. 880-3.
160. Lu, S.X., et al., *CCA1 and ELF3 Interact in the control of hypocotyl length and flowering time in Arabidopsis.* Plant Physiology, 2012. **158**(2): p. 1079-88.
161. Ni, Z., et al., *Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids.* Nature, 2009. **457**(7227): p. 327-31.
162. Farre, E.M., et al., *Overlapping and distinct roles of PRR7 and PRR9 in the Arabidopsis circadian clock.* Curr Biol, 2005. **15**(1): p. 47-54.
163. Dong, M.A., E.M. Farre, and M.F. Thomashow, *CIRCADIAN CLOCK-ASSOCIATED 1 and LATE ELONGATED HYPOCOTYL regulate expression of the C-REPEAT BINDING FACTOR (CBF) pathway in Arabidopsis.* Proc Natl Acad Sci USA, 2011. **108**(17): p. 7241-6.

164. Green, R.M. and E.M. Tobin, *Loss of the circadian clock-associated protein 1 in Arabidopsis results in altered clock-regulated gene expression.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 4176-9.
165. Lu, S.X., et al., *CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of Arabidopsis.* Plant Physiol, 2009. **150**(2): p. 834-43.
166. Seo, P.J., et al., *A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in Arabidopsis.* Plant Cell, 2012. **24**(6): p. 2427-42.
167. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-45.
168. Turner, B.M., *Cellular memory and the histone code.* Cell, 2002. **111**(3): p. 285-291.
169. Solomon, M.J., P.L. Larsen, and A. Varshavsky, *Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene.* Cell, 1988. **53**(6): p. 937-947.
170. Chaya, D. and K.S. Zaret, *Sequential chromatin immunoprecipitation from animal tissues.* Methods Enzymol, 2004. **376**: p. 361-72.
171. Ezhkova, E. and W.P. Tansey, *Chromatin immunoprecipitation to study protein-DNA interactions in budding yeast.* Methods Mol Biol, 2006. **313**: p. 225-44.
172. Sandmann, T., J.S. Jakobsen, and E.E. Furlong, *ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos.* Nat Protoc, 2006. **1**(6): p. 2839-55.
173. Ascenzi, R. and J.S. Gantt, *Subnuclear distribution of the entire complement of linker histone variants in Arabidopsis thaliana.* Chromosoma, 1999. **108**(6): p. 345-355.
174. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-837.
175. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-560.
176. Paterson, A.H., et al., *The Sorghum bicolor genome and the diversification of grasses.* Nature, 2009. **457**(7229): p. 551-556.
177. Zhang, T., et al., *Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement.* Nat Biotechnol, 2015. **33**(5): p. 531-537.
178. Wang, X., et al., *The genome of the mesopolyploid crop species Brassica rapa.* Nat Genet, 2011. **43**(10): p. 1035-1039.
179. Mayer, K.F., et al., *A physical, genetic and functional sequence assembly of the barley genome.* Nature, 2012. **491**(7426): p. 711-6.
180. Ferrier, T., et al., *Arabidopsis paves the way: genomic and network analyses in crops.* Curr Opin Biotechnol, 2011. **22**(2): p. 260-270.
181. Deng, W., et al., *Direct links between the vernalization response and other key traits of cereal crops.* Nat Commun, 2015. **6**: p. 5882.

182. Bolduc, N., et al., *Unraveling the KNOTTED1 regulatory network in maize meristems.* Genes Dev, 2012. **26**(15): p. 1685-1690.
183. Carpita, N.C., *Pectic polysaccharides of maize coleoptiles and proso millet cells in liquid culture.* Phytochemistry, 1989. **28**(1): p. 121-125.
184. Carpita, N.C., *Structure and biogenesis of the cell walls of grasses.* Annu Rev Plant Physiol Plant Mol Biol, 1996. **47**: p. 445-476.
185. Kaufmann, K., et al., *Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP).* Nat Protoc, 2010. **5**(3): p. 457-472.
186. Saleh, A., R. Alvarez-Venegas, and Z. Avramova, *An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in Arabidopsis plants.* Nat Protoc, 2008. **3**(6): p. 1018-1025.
187. Haring, M., et al., *Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization.* Plant Methods, 2007. **3**: p. 11.
188. Li, W., et al., *A robust chromatin immunoprecipitation protocol for studying transcription factor-DNA interactions and histone modifications in wood-forming tissue.* Nat Protoc, 2014. **9**(9): p. 2180-2193.
189. O'Neill, L.P. and B.M. Turner, *Immunoprecipitation of native chromatin: NChIP.* Methods, 2003. **31**(1): p. 76-82.
190. Shi, J. and J. Lai, *Patterns of genomic changes with crop domestication and breeding.* Curr Opin Plant Biol, 2015. **24**: p. 47-53.
191. Landt, S.G., et al., *ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia.* Genome Res, 2012. **22**(9): p. 1813-1831.
192. Das, P.M., et al., *Chromatin immunoprecipitation assay.* Biotechniques, 2004. **37**(6): p. 961-9.
193. Nagaki, K., et al., *Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres.* Genetics, 2003. **163**(3): p. 1221-5.
194. Bailey, T., et al., *Practical guidelines for the comprehensive analysis of ChIP-seq data.* PLoS Comput Biol, 2013. **9**(11): p. e1003326.

Vita

Dae Kwan Ko was born in Seoul, Republic of Korea on October 02, 1981. After graduating from Dongwon High School, he entered Konkuk University in Seoul, Republic of Korea. He received the degrees of Bachelor of Science in 2006 from Konkuk University. Subsequently, he entered School of Biological Sciences at Seoul National University and received Master's degree in 2008. In August of 2009, he started the Graduate School of the University of Texas at Austin as Ph.D. student in the Plant Biology program. In August of 2016, he graduated with a Doctor of Philosophy.

Permanent email: dkko@utexas.edu

This dissertation was typed by Dae Kwan Ko.